

OPTIONS

Chemistry

for the IB Diploma

SECOND EDITION

Steve Owen

with additional
online material



Option A Materials

A1 Materials science introduction

Types of materials

Materials are the substances that the things around us are made from and include stone, metals, bone, wood etc. – the list is huge. Materials are so important that human prehistory is named in terms of the materials used – Stone Age, Bronze Age etc. For scientists, engineers, architects, doctors and so on, a knowledge of materials is critical – ‘Materials’ is the study of their properties.

There are many ways of classifying materials. Traditionally, materials are classified into four main categories (Figure A.1).

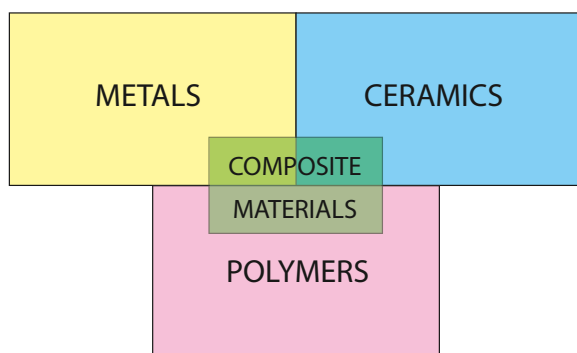


Figure A.1 A traditional way of classifying materials.

Metals

These include both pure metals and alloys – examples are iron, steel and brass (Figure A.2). Metallic elements are found on the left side of the periodic table.

Some key properties of metals are:

- good conductors of electricity
- good conductors of heat
- lustrous – shiny (when freshly cut)
- malleable – can be hammered into shape
- ductile – can be drawn into wires
- sonorous – ring when struck.

The properties of metals can be related to their structure and bonding (Topic 4). For instance, metals are good conductors of electricity because the delocalised electrons can move freely throughout the structure.

Polymers

Polymers are long-chain molecules, usually based on carbon, which are formed when smaller molecules (monomers) join together. Examples of polymers are polyethene, nylon and cellulose. The properties of polymers are many and varied and will be discussed later (page 22).

Learning objectives

- Understand that materials can be classified in different ways
- Evaluate the different ways of classifying materials
- Understand what is meant by composite materials
- Understand how bonding triangles can be constructed and used



Figure A.2 The Eiffel Tower is made from 7300 tonnes of iron.

Ceramics

Ceramics can be regarded as inorganic (not derived from plants or animals), solid engineering materials that are neither metals nor polymers. More specifically, they are usually described as inorganic substances that contain at least one metallic and one non-metallic element although substances such as silicon carbide, diamond and graphite are also usually classified as ceramics. Examples of ceramics include aluminium oxide, concrete, silicon nitride, tungsten nitride, silicon dioxide and traditional ceramics such as porcelain.

A major problem with defining ceramics is whether or not glasses should be included. These have an amorphous structure and can be regarded as supercooled liquids and therefore, according to our definition, should not be included. There are, however, some scientists who include glasses as ceramics.

Ceramics usually have the following properties:

- brittle
- hard
- strong when compressed but weaker when stretched
- resistant to chemicals
- electrical insulators – although some ceramics are superconductors
- thermal insulators.

Not all of these properties are shown by all ceramics.

Composite materials

Composite materials are mixtures that contain two or more different materials, which are present as distinct, separate phases. Synthetic composite materials consist of a reinforcing phase (e.g. carbon fibres) embedded in a matrix (e.g. a resin) (Figure A.3). Carbon fibre, steel-reinforced concrete and glass-reinforced plastic (fibreglass) are examples of composite materials.

These materials combine the ‘best’ properties of all the materials used to make the composite and they can be extremely strong and very light. Carbon fibre, for instance, is used in the construction of frames for racing bicycles – for the same strength it is lighter than steel or aluminium. Natural materials such as wood and bone are often included in the composites category.

Classification of materials

The general classification into four main groups is, of course, very broad and many other categories, such as ‘semiconductors’ and ‘biomaterials’, are often used. Each category can, of course, have several sub-categories.

There are many ways of classifying materials and different ways will be appropriate in different circumstances. For instance, an architect might need to classify materials according to properties and their suitability for a particular use; an electrical engineer might be most interested in whether or not particular materials are electrical conductors, insulators or semiconductors; a chemist would be much more interested in classification in terms of structure and bonding. No single system is perfect and the classification is chosen according to need.

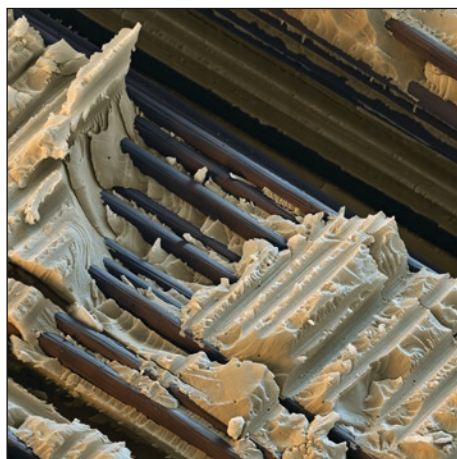
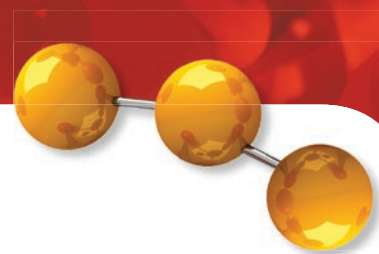


Figure A.3 A scanning electron micrograph of a carbon-fibre reinforced plastic showing the carbon fibres (reinforcing phase) embedded in the plastic (matrix) – the carbon fibres provide extra tensile strength.



According to properties

Materials can be classified in terms of a particular property, such as electrical conductivity. In this case we could classify materials as conductors (e.g. metals), insulators (e.g. diamond) or semiconductors (e.g. silicon).

Other properties that can be used to classify materials are melting point, permeability (to liquids and gases), elasticity, brittleness etc. We can generally explain properties in terms of the structure and bonding – properties such as melting point, malleability/ductility and brittleness were discussed in Topic 4.

Most materials behave elastically under certain conditions. A material exhibits elastic behaviour if, when subjected to some deforming force, it returns to its original shape and size when the force is removed. Elastic behaviour can be explained in terms of the forces between atoms/molecules/ions in a substance. When a piece of metal is subjected to a stretching force, the atoms are pulled further apart. If the metal is behaving elastically, as the force is removed the attractive forces in the lattice structure cause it to return to its original shape and size. However, if the force is too large then the planes of metal atoms slide over each other and the metal undergoes **plastic deformation**.

Permeability to moisture can be explained in terms of the type of bonding and the packing in a solid. Metals and most ceramics are generally impermeable to water because they have tightly packed structures and so there is no room for the water to pass through the structure. Certain traditional ceramics, like concrete, have porous structures and they can absorb water.

The exact nature of the water permeability of polymers depends on several factors. If the polymer is made into fibres, which are then woven into a piece of material, then permeability to water will depend on how closely woven the fabric is. For moulded plastics, the permeability to water depends on the nature of the polymer and the crystallinity. Generally, polymers that contain only carbon and hydrogen tend to have lower permeability to water because of the non-polar nature of the entire polymer chain. Polymers with a higher degree of crystallinity will have lower water permeability because the polymer chains are packed together more tightly.

According to uses

We can also classify materials according to their uses – for example biomaterials, which are materials that are used for medical implants (e.g. artificial hip joints, breast implants and contact lenses) and for other uses involving biological systems. Another category could be ‘materials suitable for use in the aerospace industry’ etc.

According to bonding

Metals have metallic bonding and polymers have covalent bonding (with London forces between chains) but the bonding in ceramics is more complicated and is a combination of ionic and covalent with the proportion of each depending on the nature of the ceramic.

A **bonding triangle** (Figure A.4) can be used to classify materials according to bonding. At each vertex of the triangle there is one of the three types of bonding.

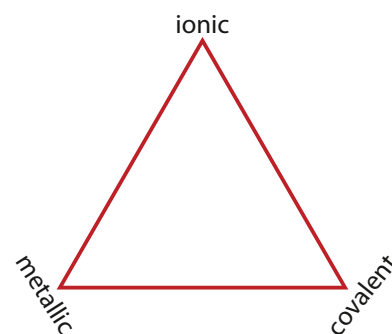


Figure A.4 A simple bonding triangle.

Binary substances (made up of just two elements) or pure elements can be arranged in bonding triangles according to their electronegativity values and the bonding between particles (Figure A.5). On the vertical axis we have the electronegativity difference between the elements and on the horizontal axis the average electronegativity.

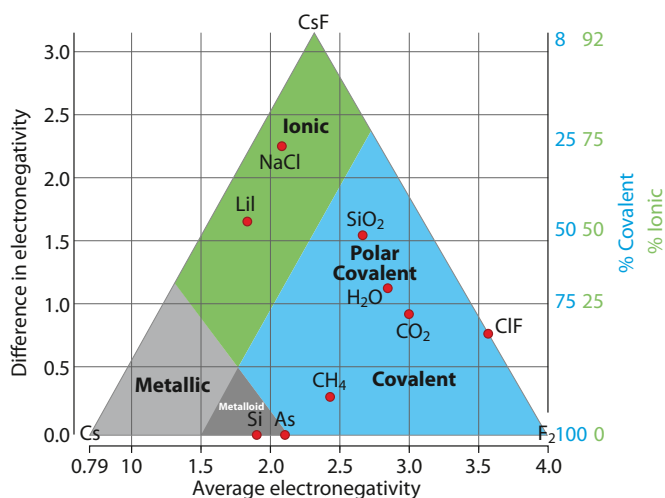


Figure A.5 A van Arkel–Ketelaar triangle showing some compounds and elements.

The triangle in Figure A.5 indicates a continuum from one type of bonding to the other two – and shows that the idea of pure ionic or pure covalent bonding is an oversimplification. Compounds can be placed in the triangle by using electronegativity values – for example, NaCl:

- Na electronegativity = 0.9
- Cl electronegativity = 3.2

$$\text{Difference in electronegativity} = 3.2 - 0.9 = 2.3$$

$$\text{Average electronegativity} = \frac{(3.2 + 0.9)}{2} = 2.1$$

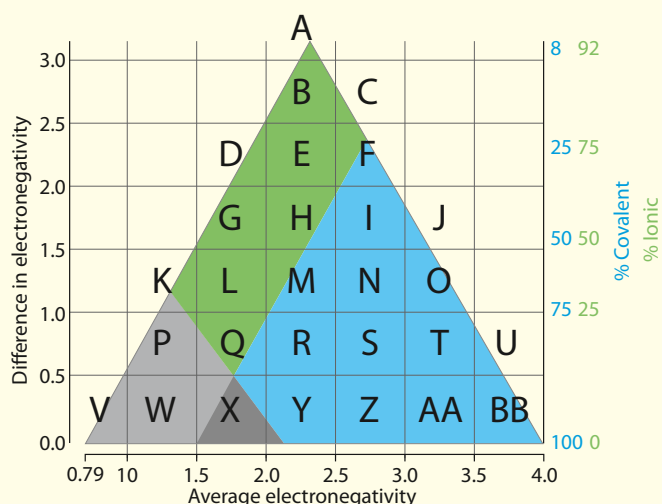
This type of bonding triangle is usually called a van Arkel–Ketelaar triangle after its originators. It is important to note that the triangle only looks at the bonding between particles and that the area marked ‘polar covalent’ in Figure A.5 refers to the **bonds** and not to the overall particle. So, CO₂ has polar bonds and occurs in this region but is a non-polar molecule overall because the dipoles cancel out.

? Test yourself

- 1 Using the electronegativity values given in the IB Chemistry data booklet, determine in which square each of the compounds below would be plotted on the bonding triangle shown below.

- | | |
|---------------------------|---------------------------|
| a GeO ₂ | d WBr ₅ |
| b ZnS | e CdS |
| c GaAs | f KCl |

State the type of bonding involved for each compound.





According to structure

We can also classify substances according to structure and this gives two broad categories – **giant** and **molecular**. Polymers generally have molecular structures with covalent bonding within molecules but weaker London forces between the polymer chains. Metals and ceramics have giant structures.

Nature of science

Science is a highly collaborative field – for example, the development of biomaterials involves collaboration between scientists from many different areas (chemists, mechanical engineers, chemical engineers, medics, vets, biologists etc.) as well as experts in various other fields such as lawyers and economists.

Many materials were used long before their properties were understood at a molecular/atomic level. As technology (electron microscopes, X-ray diffractometers etc.) has developed we have gained an insight into the structure and bonding in these materials.

A2 Metals and inductively coupled plasma spectroscopy

Extraction of metals

Metals are extracted from their ores using **redox** reactions – the metal compound (usually an oxide) is reduced to the metal either by chemical or by electrical means.

Sulfide ores are usually converted to oxides by roasting in air – the oxides are then reduced to the metal.

Different methods of extraction must be used depending on the reactivity of the metal (Table A.1).

Metal	Reactivity	Method of extraction
Na, Al	high	electrolysis
Fe, Zn, Pb	medium	heat with coke (carbon)
Cu, Ag, Au	low	found uncombined or the ore can be heated to release the metal

Table A.1 The method of extraction of a metal depends on its reactivity.

Coke, which is mostly carbon, is commonly used as a reducing agent because it is fairly cheap and readily available. Metals below carbon in the activity series can be extracted by heating with coke because carbon is a stronger reducing agent than the metal and able to ‘take the oxygen’ from the metal oxide.

Metals higher in the activity series than carbon cannot be extracted by heating with coke and electrolysis is usually used. Occasionally, reduction by a more reactive metal is used to extract a metal.

Learning objectives

- Understand that metals can be extracted from their ores by reduction with carbon or by electrolysis
- Write equations for the reduction of metal compounds to metals
- Explain the production of aluminium using electrolysis
- Solve problems involving the amount of product formed during electrolysis
- Understand what is meant by the term alloy
- Explain how alloying can change the properties of metals
- Understand what is meant by the terms diamagnetism and paramagnetism
- Explain the basic principles of inductively coupled plasma (ICP) spectroscopy
- Use data from ICP experiments to calculate the amount of a metal in a sample

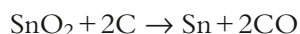
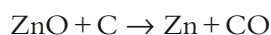
Coke is formed by heating coal in the absence of air.

Other reducing agents such as carbon monoxide are also used.

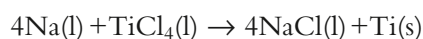
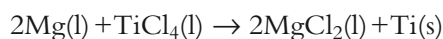
The product of these reactions is shown as carbon monoxide, which is the most favourable reaction under these conditions, but carbon dioxide is also sometimes shown as the product.

The reaction must be carried out in an argon atmosphere to prevent magnesium or sodium from reacting with oxygen in the air. Magnesium and sodium are expensive metals as they are extracted by electrolysis and the overall process (the Kroll process) for the extraction of titanium is very expensive, making the metal very expensive.

For example, zinc and tin are obtained from their respective oxides by heating with carbon:



Titanium could be extracted from titanium(IV) oxide by heating with carbon at a temperature above about 1600 K but the problem is that titanium reacts to form titanium carbide, which makes the metal brittle. It is therefore extracted by heating the chloride (TiCl_4) with sodium or magnesium:



Sodium and magnesium are stronger reducing agents (higher in the activity series) than titanium.

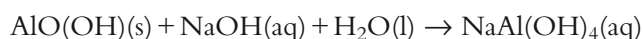
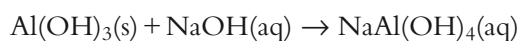


The economic recession which began in 2007/8 resulted in huge rises in the prices of certain metals. In some countries this resulted in increases in the amount of thefts of, for instance, copper cabling, lead from roofs and catalytic converters but in other countries it had more deadly consequences. The rise in gold prices resulted in many illegal gold mines springing up in Nigeria and lead dust produced in the process is believed to be responsible for the deaths, through lead poisoning, of hundreds of children in the areas surrounding the gold mines.

Extraction of aluminium

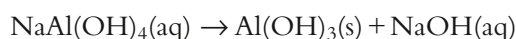
Refining bauxite

The main ore of aluminium is bauxite, which is mostly a mixture of aluminium hydroxide (Al(OH)_3) and hydrated aluminium oxides (AlO(OH)). Before electrolysis, the bauxite must be purified. The ore is crushed and then dissolved in hot sodium hydroxide at 175°C – the aluminium compounds dissolve to form sodium tetrahydroxoaluminate (NaAl(OH)_4):



The impurities are insoluble and can be filtered out.

The resulting solution is cooled and this causes solid aluminium hydroxide to precipitate from solution:



The aluminium hydroxide crystals are removed from the solution and heated to over 1000°C . This causes the hydroxide to decompose into alumina (aluminium oxide) and water:



This process, called the Bayer process, relies on the **amphoteric** nature of aluminium oxide and aluminium hydroxide and the basic nature of other metal oxides. Amphoteric oxides and hydroxides can act as a **base** or an **acid**. In the Bayer process, aluminium hydroxide is acting as an acid, reacting with sodium hydroxide.

Electrolysis of alumina



Because aluminium is more reactive than carbon, aluminium oxide cannot be reduced to aluminium by heating with carbon and electrolysis must be used.

Alumina (aluminium oxide) is an ionic solid made up of Al^{3+} and O^{2-} ions. In order to conduct electricity, the ions must be free to move. This requires melting the alumina (so that the strong electrostatic forces between the oppositely charged ions are overcome). However, alumina has a very high melting point (2072°C) because of the high charges on the individual ions. Heating alumina to this temperature, and then maintaining it for the electrolytic process, would require a lot of energy and be extremely expensive. Alumina is therefore dissolved in molten **cryolite** (sodium aluminium fluoride, Na_3AlF_6) which melts at only 1012°C . A solution of aluminium oxide in cryolite is formed and this costs much less to keep molten. Dissolving the alumina causes it to separate into positive and negative ions.

The electrolysis of dissolved alumina is called the **Hall–Héroult process** (Figure A.6). The electrolytic cell is made of steel, with a refractory ceramic lining to withstand the high temperatures. The base of the cell is further lined with graphite, which acts as the cathode. Several large graphite blocks suspended from a support act as the anode.

Graphite is used for a number of reasons:

- it conducts electricity well
- it is cheap and easily replaced
- it is relatively inert
- it has a high melting point, well above that of the molten alumina.

Hundreds of cells are lined up in series and a huge current of several hundred thousand amps is passed through the circuit. The application of a current causes the free ions to move towards their respective oppositely charged electrodes.

Smelting is an expensive and energy-intensive process that uses huge amounts of electricity. Many aluminium smelters are located close to hydroelectric or other power stations to ensure a good (and preferably cheap) supply of electricity.

The resistance to the passage of electricity through the molten cryolite generates enough heat to keep the compounds molten without the need for an external heat source such as a furnace.

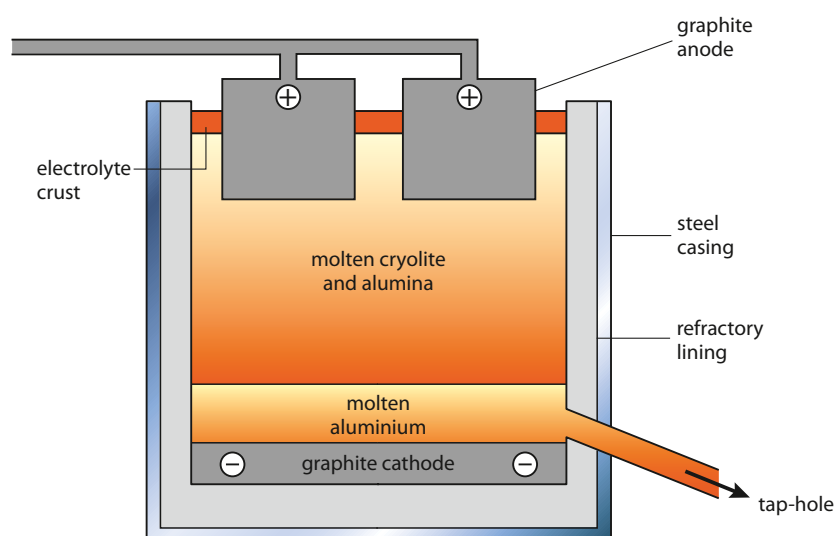
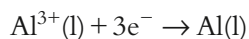


Figure A.6 The Hall–Héroult cell for the electrolysis of molten alumina/cryolite.

Exam tip

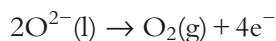
The usefulness of state symbols on the ions in these half-equations is dubious, but if one is required in the examination then '(l)' is probably most appropriate.

Aluminium ions are attracted to the cathode (negatively charged) at the base of the cell, where they gain electrons and are reduced to liquid aluminium:



Elemental aluminium is denser than molten cryolite, and so remains pooled at the bottom of the cell until it is tapped off.

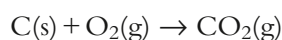
Oxide ions are attracted to the graphite anodes and are oxidised to oxygen.



The overall reaction for the process is:



The high temperature in the cell causes the oxygen gas produced at the anode to react with the graphite, oxidising the carbon to carbon dioxide:



The anodes gradually erode away and have to be replaced periodically.

Quantitative electrolysis

We can work out the amount of metal produced when a molten salt is electrolysed by using an approach similar to that used in mole calculations in Topic 1.

Current is the amount of charge that passes a certain point per second so by knowing the current and the time of electrolysis we can work out the amount of charge that passes. The relationship between charge, current and time is:

$$Q = I \times t$$

where Q is the charge in coulombs, I is the current in amperes and t is the time in seconds.

So, if a current of 2.00 A is passed for 1.00 hour, the charge that flows is:

$$Q = 2.00 \times 1.00 \times 60 \times 60 = 7200 \text{ C}$$

To work out the number of moles of metal produced we need to know how many electrons flow around the circuit. The charge on an electron is $1.6 \times 10^{-19} \text{ C}$, therefore the charge on one mole of electrons is approximately 96 500 C.

If we divide the charge that flows around the circuit by 96 500, we will get the number of moles of electrons that flow around the circuit, from which we can work out the number of moles of metal produced.

96 500 C mol⁻¹ is known as the **Faraday constant** and represents the charge on one mole of electrons.



Worked example

A.1 In the electrolysis of molten calcium chloride, a current of $5.00 \times 10^2 \text{ A}$ is passed for 10.0 hours. Calculate the mass of calcium formed.

The time must be converted to seconds: $10.0 \times 60 \times 60 = 36\,000 \text{ s}$

The charge that flows can be worked out using $Q = I \times t$

$$= 500 \times 36\,000$$

$$= 1.80 \times 10^7 \text{ C}$$

The number of moles of electrons can be worked out by dividing the charge by the Faraday constant:

$$\begin{aligned}\text{number of moles of electrons} &= \frac{1.80 \times 10^7}{96\,500} \\ &= 186.5 \text{ mol}\end{aligned}$$

More significant figures will be carried through on the calculator to the next part of the question.

The half-equation for the reduction of calcium ions at the cathode is: $\text{Ca}^{2+} + 2\text{e}^- \rightarrow \text{Ca}$

It can be seen from this that two moles of electrons are required to produce one mole of calcium.

So to work out the number of moles of calcium formed we must divide the number of moles of electrons by 2:

$$\begin{aligned}\text{number of moles of calcium} &= \frac{186.5}{2} \\ &= 93.26 \text{ mol}\end{aligned}$$

The mass of calcium produced can be worked out by multiplying the number of moles by the relative atomic mass of calcium:

$$\begin{aligned}\text{mass of calcium} &= 93.26 \times 40.08 \\ &= 3740 \text{ g}\end{aligned}$$

The final answer is given to three significant figures, which is consistent with the data in the question.

We can summarise the steps in working out an electrolysis problems as:

- 1 Calculate the amount of charge that flows using $Q = I \times t$ (remember that current must be in amps and the time in seconds).
- 2 Divide the charge by 96 500 to give the number of moles of electrons.
- 3 Write the half-equation for the reaction at the electrode to produce one mole of product.
- 4 Divide the number of moles of electrons by the coefficient of the electrons in the half-equation. This gives the number of moles of product formed.
- 5 Convert the number of moles of product formed to a mass by multiplying by the relative atomic mass (or relative molecular mass for gases).

? Test yourself

- 2 Use the activity series in the IB Chemistry data booklet to work out if the following metals could be extracted from their ores by heating with coke:
 - a sodium
 - b lead
 - c strontium
 - d cadmium
- 3 Write equations for the reduction of these metal oxides using coke:
 - a Bi_2O_3
 - b CuO
 - c Fe_2O_3
- 4 Write overall equations for the electrolysis of these molten salts:
 - a MgCl_2
 - b KCl
- 5 Calculate the amount of charge that flows in each of the following:
 - a a current of 5.00 A flows for 100 seconds.
 - b a current of 8.00 A flows for 3.00 hours.
 - c a current of 4.00 mA flows for 5.25 hours.
- 6 Calculate the number of moles of electrons that pass when:
 - a a current of 2.20 A flows for 900 seconds.
 - b a current of 100 A flows for 6.00 hours.
 - c a current of 200 mA flows for 24.0 hours.
- 7 Work out the mass of metal formed in each of these electrolyses:
 - a a current of 50.0 A is passed through molten lithium chloride for 10.0 minutes.
 - b a current of 10.0 A is passed through molten magnesium chloride for 2.00 hours.
 - c a current of 20.0 A is passed through molten aluminium oxide for 15.0 hours.

Alloys and the magnetic properties of metal compounds

Alloys

Alloys are homogeneous mixtures of two or more metals, or of a metal with a non-metal.

The majority of metals that we come across in everyday life are alloys, rather than pure metals. Steel, an alloy of iron and carbon, is used much more extensively than pure iron. The properties of some different forms of steel are shown in Table A.2.

Type of steel	Composition and properties	Uses
mild steel	0.15–0.3% carbon, cheap, malleable, not ductile, will corrode	structural steel used in construction
medium carbon steel	0.3–0.5% carbon, wear resistant, balanced strength and ductility	car parts (body and engine)
high/ultra-high carbon steel	0.5–2% carbon, very strong and hard	springs, high-strength wires, specialist uses, e.g. punches, axles
stainless steel	15% chromium, 10% nickel, corrosion resistant, strong, hard	cutlery, kitchenware, surgical equipment, major appliances

Table A.2 The properties and uses of different forms of steel.

Alloys tend to be stronger and stiffer than pure metals and often combine the desirable properties of the different metals involved. For example, aluminium is a light (low density) metal but it is not strong enough to be used in aeroplane manufacture until it is alloyed with copper (and smaller amounts of magnesium and manganese) to produce duralumin.

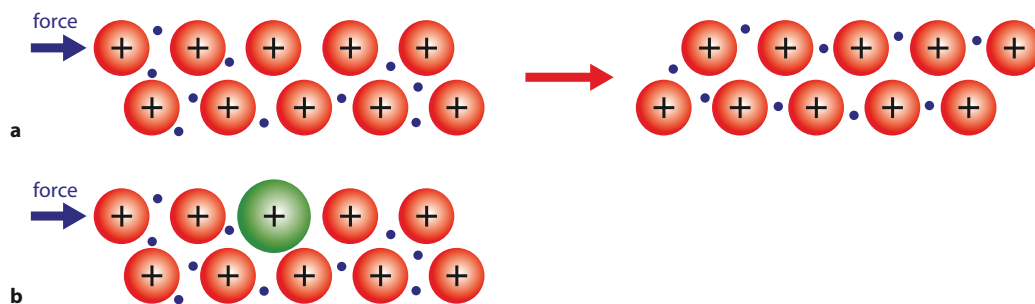


Figure A.7 **a** Metals are malleable and ductile because the planes of atoms/ions can slide over each other without disrupting the bonding. **b** The introduction of a larger atom makes it more difficult for the planes of atom/ions to slide over each other and so alloys tend to be stronger and stiffer than pure metals.

The reason that alloys are stronger than the pure metals can be explained in terms of the structure of metals. At the simplest level, we can imagine that different-sized atoms will prevent planes of metal atoms sliding over each other as easily (Figure A.7b).

Not all properties of alloys are desirable – for example:

- aluminium alloys are more susceptible to corrosion than the pure metal
- the electrical conductivity of copper is reduced by alloying with other metals.

Magnetic properties of metal compounds

There are two forms of magnetism that we need to be concerned with here and these are **paramagnetism** and **diamagnetism**.

All substances have some paired electrons and so all substances exhibit diamagnetism, but the diamagnetic effect is much smaller than the paramagnetic effect and so, if there are any unpaired electrons present, the paramagnetic effect will dominate and the substance will be paramagnetic overall and attracted by a magnetic field. The more unpaired electrons that are present, the greater the paramagnetism (magnetic moment).

Consider the electronic configurations of two transition metal ions shown in Figure A.8. Both ions contain unpaired electrons and so compounds containing them, such as FeCl_2 , will be paramagnetic. Because an Fe^{2+} ion has four unpaired electrons, but the Cr^{3+} ion has only three, iron(II) compounds are more paramagnetic (have a higher magnetic moment) than chromium(III) compounds.

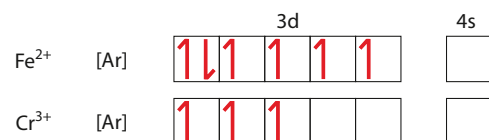


Figure A.8 Electronic configurations of Fe^{2+} and Cr^{3+} .

The Cu^+ ion has the electronic configuration shown in Figure A.9. Because all the electrons are paired, compounds of copper(I), such as CuCl , are diamagnetic.



Figure A.9 Electronic configuration of Cu^+ .

Extension

Actually metals are more ductile than this simple picture of planes sliding over each other would predict and this can be better explained by the movement of dislocations through a metallic lattice. Dislocations are imperfections in the lattice structure and can allow the planes to move relative to each other more easily.

- **paramagnetism is caused by the presence of unpaired electrons and paramagnetic substances are attracted by a magnetic field.**
- **diamagnetism is caused by the presence of paired electrons and diamagnetic substances are repelled slightly by a magnetic field.**

In the absence of an external magnetic field the spins of the unpaired electrons in a paramagnetic substance are arranged randomly and there is no magnetic effect. However, when a magnetic field is applied the spins of the unpaired electrons align with the magnetic field causing a paramagnetic effect. Substances such as iron are ferromagnetic and the unpaired electrons are aligned even in the absence of a magnetic field.

Extension

The situation is more complicated with complex ions because, depending on the energy difference between the higher and lower sets of d orbitals and the amount of energy required to pair up two electrons in the same d orbital (overcoming the repulsions), the complexes may be high spin (maximum number of unpaired electrons) or low spin (maximum number of electrons in the lower set of d orbitals). Just how paramagnetic the substance is then depends on the ligands, which influence the splitting of d orbitals.

The magnetic behaviour of metals themselves is much more complicated than this and an explanation requires a more advanced treatment of the bonding in terms of band theory. Most metals are weakly paramagnetic due to an effect called Pauli paramagnetism where, in the presence of an applied magnetic field, it is more favourable for some electrons to be promoted to a slightly higher energy level so that their spins can be aligned with the applied field (lower energy state). This means that there is an excess of unpaired electrons and the metal is slightly paramagnetic. This effect is generally much smaller than the paramagnetism due to unpaired electrons in transition metal compounds. It is difficult to draw general conclusions about the paramagnetism of metals but generally transition metals, lanthanoids and actinoids tend to be more paramagnetic than metals in Group 1 and 2 of the periodic table. So, the presence of electrons in d and f orbitals influences the magnetic properties. Copper and zinc (and the other elements in groups 11 and 12), which have full d subshells are diamagnetic.

Inductively coupled plasma detection techniques

Inductively coupled plasma (ICP) techniques can be used to identify the presence of and determine the amount/concentration of trace (very small) amounts of metal (and some non-metal) atoms/ions present in a sample. There are two main variations on the technique – they are called ICP-OES and ICP-MS.

They have applications in the food industry (analysing for contaminants, such as mercury in shell fish), in analysing biological samples (e.g. lead in tissue samples or the ratio between ^{235}U and ^{238}U in urine using ICP-MS), in geology (e.g. determination of the amount of lanthanum in a mineral sample using ICP-OES), environmental science (e.g. analysis of cadmium in water/soil) etc. and can detect concentrations at the $\mu\text{g dm}^{-3}$ level (parts per billion).

Plasma is the fourth state of matter – in addition to solid, liquid and gas.

A plasma is a fully or partially ionised gas consisting of positive ions and electrons – the plasma is usually electrically neutral overall.

The electrostatic interactions between the charged particles in a plasma give it special properties that are very different from the properties of a gas. Because of the charged particles present in a plasma, it is a good conductor of electricity and can interact with both electric and magnetic fields.

Inductively coupled plasma-optical emission spectroscopy (ICP-OES)

This technique involves the generation of a plasma containing the sample. The plasma is generated using argon gas. A spark (from a Tesla coil) causes some of the argon atoms to become ionised. An oscillating radio frequency induction coil causes the electrons to move back and forth in a circular path and collide with other argon atoms generating more charged particles. This movement and the collisions of charged particles due to coupling with the induction coil generates heat so that temperatures of up to 10 000 K can be reached in the plasma.

ICP-OES is also called 'inductively coupled plasma-atomic emission spectroscopy' (ICP-AES)

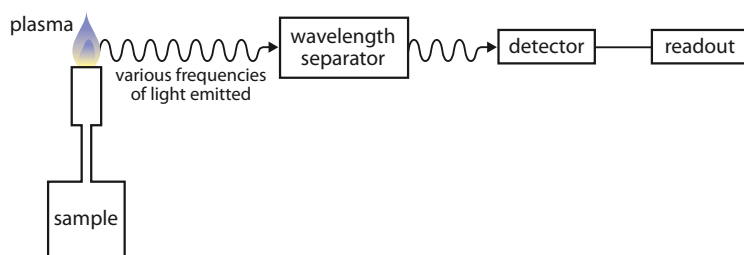


Figure A.10 Simple block diagram of an ICP-OES instrument.

The sample to be analysed is introduced into the flow of argon gas and when it enters the plasma atoms/ions are produced in an excited state. This means that there are electrons in higher energy levels than the normal ground state and subsequently light will be given out as the electrons fall back down to lower energy levels (see atomic emission spectra in Topic 2). We are basically generating the emission spectrum of the atoms/ions. The frequencies of the light produced are characteristic of a particular element, and the intensity of a particular frequency of light emitted (line in the spectrum) is related to the amount of that element present in the sample.

The instruments can analyse for several elements at the same time by separating the wavelengths of light emitted using a diffraction grating or a prism. The different wavelengths enter a photomultiplier tube which produces an electrical signal, the magnitude of which depends on the number of photons entering the tube (Figure A.10).

In order to measure the amount/concentration of a particular metal present in a sample, the instrument must be calibrated by using samples of known concentration of a particular atom/ion. The intensity of one particular line (one wavelength) in the emission spectrum is measured for each concentration and then a calibration curve is plotted (Figure A.11). When an unknown sample is introduced into the instrument we can read off the concentration from the curve. For instance, if the intensity of a particular spectral line emitted by our unknown sample is 54 (arbitrary units), we can read off the concentration of calcium in the sample as 6.8 mg dm^{-3} .

Note on units: mg dm^{-3} and $\mu\text{g cm}^{-3}$ are equivalent units.
 $10 \text{ mg dm}^{-3} = 10 \mu\text{g cm}^{-3}$.
 These units are also usually regarded (see Topic 1) as being equivalent to parts per million, so that
 $10 \text{ mg dm}^{-3} = 10 \mu\text{g cm}^{-3} = 10 \text{ ppm}$

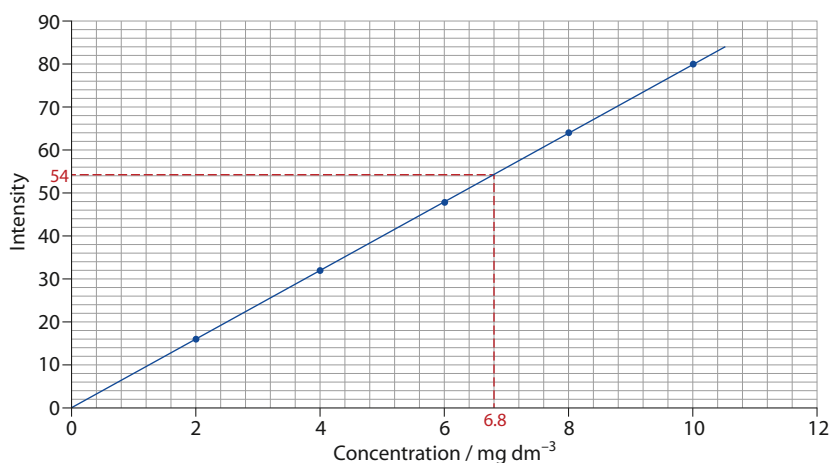


Figure A.11 Calibration curve for calcium.

? Test yourself

- 8 A sample of mineral water was analysed for calcium by ICP–OES. The emission reading was 74. Use the calibration curve in Figure A.11 to determine the concentration of calcium ions in the water in mg dm^{-3} and mol dm^{-3} .

Inductively coupled plasma–mass spectrometry (ICP–MS)

Mass spectrometry can also be used to analyse the ions formed in a plasma. Positive ions from the plasma are fed into the mass spectrometer via an interface (Figure A.12) that allows the pressure to be reduced (mass spectrometers operate under vacuum to prevent collisions with air molecules).

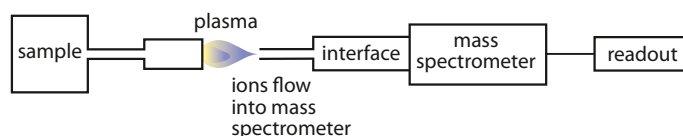


Figure A.12 Simple block diagram of an ICP–MS instrument.

Ions are separated in a mass spectrometer according to their mass:charge (m/z or m/e) ratio (essentially the same as mass for a singly charged ion). A basic diagram of a magnetic sector mass spectrometer used in some ICP–MS instruments is shown in Figure A.13. The ions are accelerated using an electric field and then deflected in a magnetic field. Ions with a small mass:charge (m/z) ratio are deflected more and the magnetic field strength is changed to bring ions of each mass to the detector in turn. For one particular magnetic field, particles of only one m/z ratio pass through the spectrometer. These hit the detector and produce a signal in the form of an electric current which is proportional to the number of ions hitting the detector.

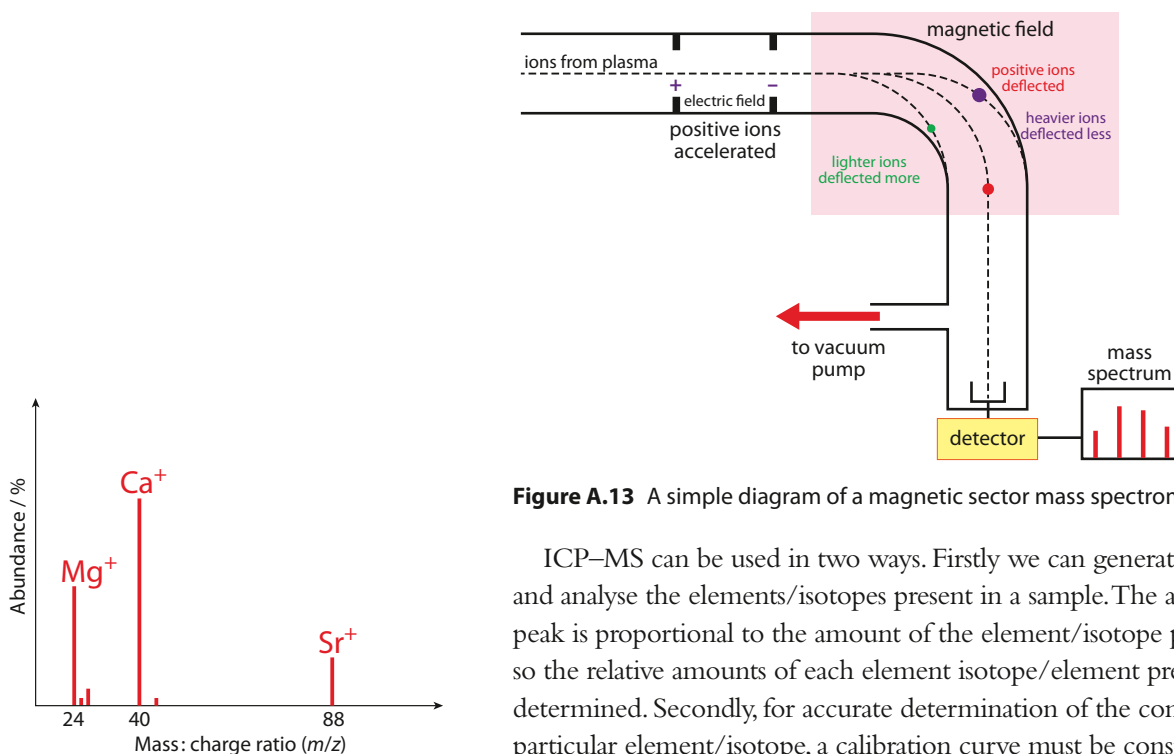


Figure A.14 The mass spectrum produced by ICP–MS of a sample containing some group 2 metals.

Figure A.13 A simple diagram of a magnetic sector mass spectrometer.

ICP–MS can be used in two ways. Firstly we can generate a mass spectrum and analyse the elements/isotopes present in a sample. The area under each peak is proportional to the amount of the element/isotope present – and so the relative amounts of each element isotope/element present can be determined. Secondly, for accurate determination of the concentration of a particular element/isotope, a calibration curve must be constructed for each isotope present in the sample by using known concentrations.

Figure A.14 shows the mass spectrum of a sample containing some group 2 elements.



Nature of science

Developments in science and technology often go hand-in-hand. For example, the development of alloys with enhanced properties has allowed the design of aeroplanes that can fly faster and higher.

Scientific knowledge is constantly increasing and this can have many benefits for society. The development of ICP techniques has allowed the determination of trace amounts of metals, that could be harmful, in samples of soil/food/tissue.

A3 Catalysts

Many spontaneous chemical and biological reactions occur incredibly slowly at room and body temperatures. Catalysts are of vital importance to manufacturing and life processes because they increase the rate of these reactions without the need for dramatic changes to the reaction conditions. Catalysts are unchanged at the end of the reaction.

Catalysts provide an alternative reaction pathway that has a lower activation energy than the uncatalysed pathway.

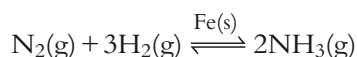
Types of catalysts

Catalysts can be broadly categorised as **homogeneous** or **heterogeneous**, depending on the phase in which they and the reactants exist.

Heterogeneous catalysts

A **heterogeneous catalyst** is one in a different phase (state) from the reactants.

Heterogeneous catalysts are usually solids and reactions occur on their surface. Transition metals and their compounds are particularly good at adsorbing (note '**ads**orbing', not '**abs**orbing') gases, and so they are commonly used as heterogeneous catalysts in industry – for example iron is used in the Haber process for the production of ammonia and nickel in the hydrogenation of unsaturated hydrocarbons. In the iron-catalysed production of ammonia, the reactants are gases but the catalyst is a solid:



Heterogeneous catalysts rely on their ability to **adsorb** reactant molecules onto active sites on their surfaces (Figure A.15). These active sites are sites on the surface that are better able to catalyse the reaction (due to structural and/or electronic factors).

Learning objectives

- Understand the differences between heterogeneous and homogeneous catalysis
- Understand the role of nanoparticles in catalysis
- Understand that zeolites can be used as selective catalysts
- Discuss the factors that influence the choice of catalysts in industry

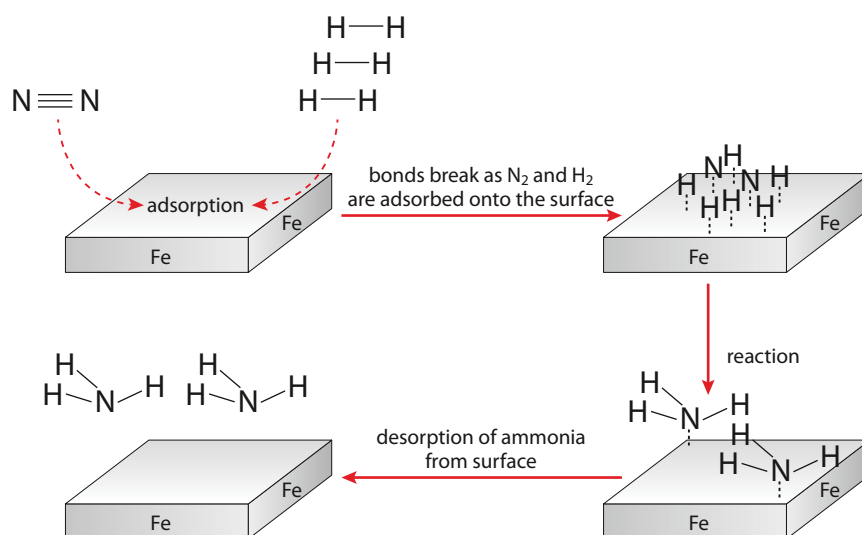


Figure A.15 The reaction of nitrogen and hydrogen on an iron surface.

Exam tip

Adsorption means that the reactants bind to the **surface** of the catalyst – it is not the same as absorption.

Adsorption increases the localised concentrations of the reactants (because they are held on the surface of the catalyst) and thereby increases the collision rate – it can also bring the reactants together in the correct orientation for reaction. Also, at the active sites on the surface, covalent bonds in the reactants are weakened/broken and this reduces the activation energy barrier for the reaction.

Many heterogeneous catalysts are added to reaction mixtures in powder form, as a fine mesh or attached to structures that have a large surface area (e.g. in catalytic converters for motor vehicles). This is because heterogeneous catalysis occurs only at the surface of the catalyst. Many industrial reactions now employ nanoparticles (particles with a diameter of 100 nm or less) on a porous support as heterogeneous catalysts. These have an extremely high surface area per unit mass and, therefore, a very large number of active sites are available for reaction. The particles need to be supported in some way to allow easier removal from the reaction mixture and to prevent aggregation of the particles.

Carbon nanocatalysts

There is currently much research looking at the possibility of using carbon nanotubes (see page 31) as both supports for catalysts and as catalysts themselves. For instance, carbon nanotubes could replace the much more expensive platinum as the catalyst in some applications such as fuel cells. Carbon nanotubes are useful as heterogeneous catalysts because they have an extremely high surface area and can coordinate other atoms and groups of atoms.

Zeolites

Zeolites are aluminosilicate (composed mostly of Si, Al and O) structures that have a **cage structure** containing a large number of pores – channels through the structure and cavities (Figure A.16). In naturally occurring zeolites, these pore sizes are up to about 1 nm, which can be compared to the diameter of a molecule of benzene (C_6H_6), which is about 0.6 nm. This means that zeolites can provide a very large surface area for catalytic reactions due to the large amount of internal surface available to reactant molecules. Zeolites can have surface areas up to about 800 m² per gram.

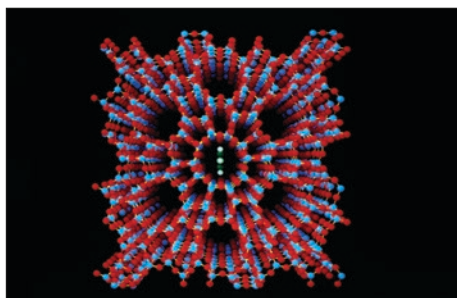


Figure A.16 The structure of a zeolite showing channels through the structure.

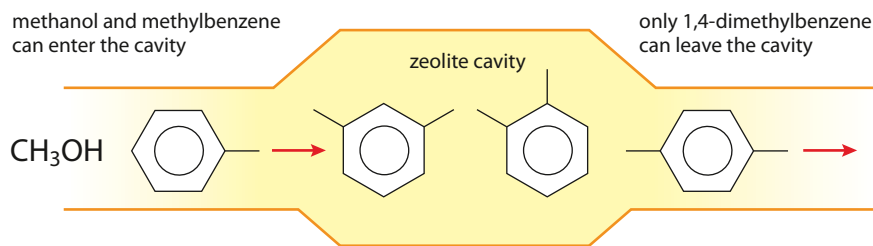


Figure A.17 Selective catalysis.

Zeolites can act as size- and shape-selective catalysts because of this pore structure. This can work in various ways – for instance, only reactant molecules of less than a certain size can fit into the channels and reach the majority of the catalyst active sites. In the reaction between methanol and methylbenzene, three possible isomers can be formed but we can select for just one product by using a zeolite catalyst. If the reaction occurs in a cavity in the zeolite (Figure A.17) only 1,4-dimethylbenzene is able to escape from the cavity and the other forms undergo isomerisation to produce more of this.

Homogeneous catalysts

The most common form of homogeneous catalysis involves having both the catalyst and the reactants in solution (aqueous or organic). Enzymes are homogeneous catalysts which are important in biological processes.

A homogeneous catalyst is in the same phase (state) as the reactants.

Homogeneous catalysts usually work by enabling a reaction to occur by a different mechanism from the uncatalysed mechanism. This involves the catalyst forming an intermediate with one or other of the reactants. For example, if C is a catalyst in the reaction between A and B, the uncatalysed reaction is:



... and the catalysed reaction is:



E_1 is bigger than either of activation energies E_2 or E_3 (Figure A.18). The catalyst is reformed at the end of the reaction.

A catalyst can also work by forming a temporary interaction with the transition state, which stabilises it and therefore lowers the activation energy – this is how enzymes catalyse reactions (Figure A.19).

Transition metal compounds often act as homogeneous catalysts. The ability to act as a catalyst relies on a transition metal atom/ion being able to exhibit various oxidation numbers and also to coordinate other molecules and ions.

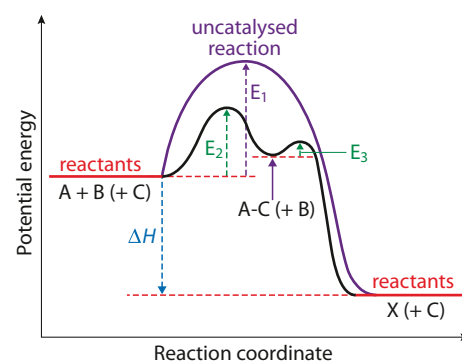


Figure A.18 The formation of an intermediate in a homogeneous catalysis reaction.

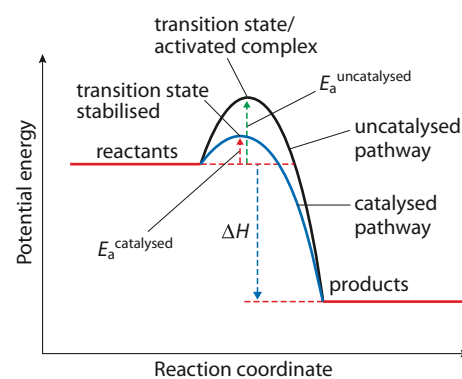
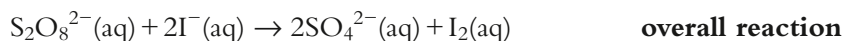
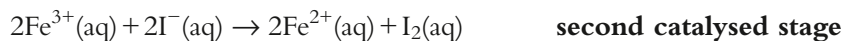
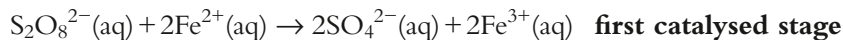


Figure A.19 An interaction between a catalyst and a transition state can lower the activation energy.

An example of homogeneous catalysis is the iron(II)–catalysed reaction between persulfate ions ($\text{S}_2\text{O}_8^{2-}$) and iodide ions:



The reaction occurs in two steps – in the first step the Fe^{2+} ion is oxidised to Fe^{3+} and then, in the second step, it is reduced back to Fe^{2+} :



The Fe^{2+} ion is regenerated in the second stage, so overall it has not been used up.

The coordination of molecules around a transition metal ion can be seen in the mechanism for the formation of high-density polyethene using a Ziegler–Natta catalyst (Figure A.20).

Besides transition metal ions, another common homogeneous catalyst is the proton (H^{+}). Reactions in which H^{+} is the catalyst are called **acid-catalysed** reactions. Carboxylic acids react with alcohols to form esters in an acid-catalysed reaction.

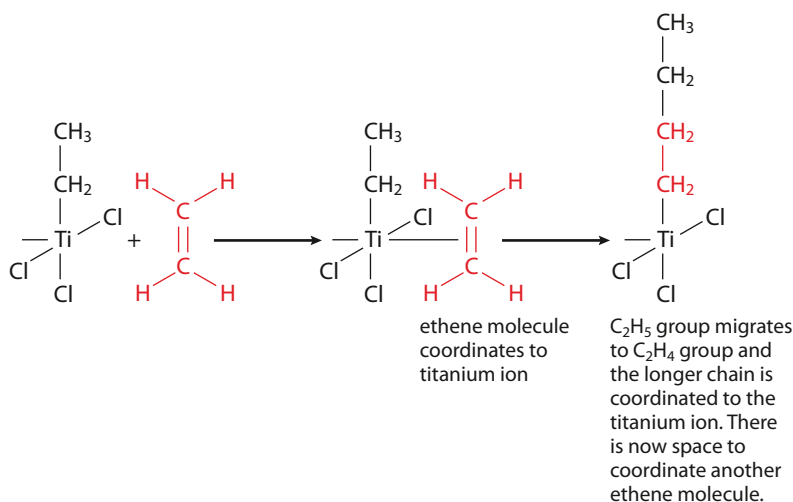


Figure A.20 Catalysing the polymerisation of ethene.

Choice of catalytic method

Over 90% of industrial processes use heterogeneous catalysis, despite many advantages of homogeneous catalysis. This is almost entirely down to the ease of separating a heterogeneous catalyst from the reaction products. But what factors should an industrial chemist take into account when deciding which catalyst to use for a given reaction?

How specific is the catalyst?

In other words, does the catalyst only catalyse one particular reaction? Homogeneous catalysis is far more specific to a particular reaction. If selectivity of the product is desired, homogeneous catalysis is more useful. Supported enzymes are also highly selective as catalysts and we have seen above how zeolites (heterogeneous catalysts) can be used in this way.



How efficient is the catalyst?

How fast is the catalysed reaction and what degree of product conversion (yield) is obtained? Chemists must consider whether efficiency is essential or if the unreacted material can be recycled. Homogeneous catalysis is more efficient because of higher availability of active sites. Only certain sites on the surface of a heterogeneous catalyst are usable – atoms in the bulk of the structure are not available – but with some homogeneous catalysts every atom or ion is available.

Does the reaction require severe conditions?

Homogeneous catalysts tend not to work well in extreme conditions (such as high temperature) whereas heterogeneous transition metal catalysts withstand high temperatures and pressures well. Some heterogeneous catalysts work only if the temperature is high enough – high temperatures and pressures can be expensive to generate and maintain, and can also affect the yield of reactions.

How is the catalyst affected by impurities in the reaction mixture?

Both heterogeneous and homogeneous catalysts can be poisoned by impurities in the reaction mixture. Heterogeneous catalysts become poisoned by the build-up of substances such as sulfur or carbon on their surface. It is sometimes possible to regenerate poisoned heterogeneous catalysts. Once a homogeneous catalyst has been deactivated, it generally needs to be replaced completely.

Are there any environmental considerations in the use and disposal of the catalyst?

For example the disposal of heavy metal catalysts, or catalysts that have been poisoned by heavy metals, can cause environmental problems.

Leaded gasoline cannot be used in cars with catalytic converters because the lead poisons the heterogeneous catalyst.

Nature of science

There is no all-encompassing scientific method. Finding a suitable catalyst for a reaction is often a matter of trial and error – and serendipity can play an important role.

It is often not essential to understand how a catalyst works to develop its use for a particular process. However, as our understanding of chemical processes increases, ever more sophisticated models of how catalysts work are being developed. This means that some of the trial and error is being taken out of the process of finding a suitable catalyst.

Learning objectives

- Understand what is meant by a liquid crystal
- Describe the behaviour of thermotropic and lyotropic liquid crystals
- Explain the functioning of a liquid crystal in terms of the arrangement of the molecules
- Understand what is meant by a nematic liquid crystal phase
- Understand what properties are suitable for a liquid crystal to be used in a liquid crystal display (LCD)

A4 Liquid crystals

What is a liquid crystal?

A 'liquid crystal' is a state of matter in which the properties of a compound exhibit some of the characteristics of both a liquid and a solid. They are fluids with electrical, optical, elastic and some other physical properties that depend on the orientation of their molecules relative to some fixed axis in the material. Generally, in the liquid crystal phase, the molecules are orientated uniformly (point in the same direction) as in a solid crystal, but retain the ability to flow and move as in a liquid.

Examples of substances that possess liquid crystal properties under certain conditions include cellulose, DNA and the solution secreted by spiders to make their silk. These substances do not necessarily display liquid crystal properties in their standard states.

Thermotropic liquid crystals

Materials that show thermotropic properties are pure substances that exist in the liquid crystal phase over only a certain **temperature** range between the true solid and liquid phases (Figure A.21).

If the temperature rises too high, the molecular orientation is disrupted because the molecules gain kinetic energy and a liquid forms. Too low a temperature causes the substance to form a normal solid crystal with no fluid properties. The **biphenyl nitriles** used in some liquid crystal displays are examples of thermotropic liquid crystals.

4-cyano-4'-pentylbiphenyl (Figure A.22) is a liquid crystal between 18 and 36 °C, giving it liquid crystal properties at room temperature.

The molecule in Figure A.22 (and other biphenyl nitriles) can be roughly described as **rod-shaped** – it is significantly longer in one direction than the other two. In the liquid crystal phase, these rods show some degree of alignment with the molecules, on average, pointing in the same direction (Figure A.23). However, there is no specific positional order – the molecules are positioned randomly relative to each other, so they can flow past each other. This phase is called the **nematic** phase.

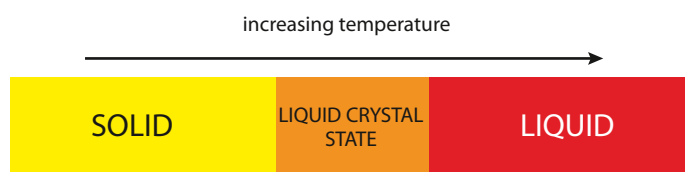


Figure A.21 The changes that occur as a substance with thermotropic liquid crystal properties is heated.

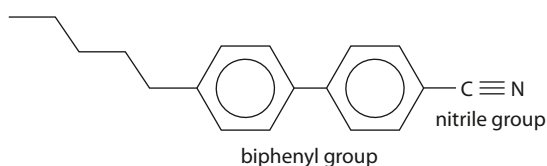


Figure A.22 The liquid crystal molecule 4-cyano-4'-pentylbiphenyl – a member of the biphenyl nitriles.

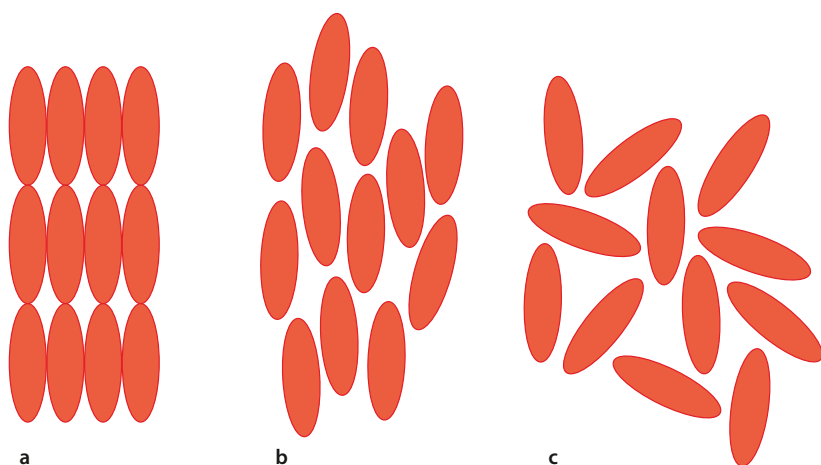


Figure A.23 The orientation and distribution of molecules in a thermotropic liquid crystal: **a** solid – regular arrangement and orientation; **b** nematic liquid crystal phase – random arrangement but fairly regular orientation; **c** above the liquid crystal temperature range – random orientation and arrangement.

Lyotropic liquid crystals

These are formed by materials depending on the **concentration** of the compound in solution. The detergent molecules in soap show lyotropic properties. The stearate ion in sodium stearate ($\text{C}_{17}\text{H}_{35}\text{COO}^-\text{Na}^+$), a constituent of many soaps, contains a long hydrophobic, non-polar hydrocarbon chain and a polar, hydrophilic carboxylate (COO^-) group. In a dilute aqueous solution, the distance between molecules is relatively large, and they do not show any order in their orientation. However, when the concentration increases, the molecules begin to line up in a specific manner in order to minimise the interactions between the hydrophobic chains and the water molecules. Soap molecules and related compounds, such as the phospholipids found in cell membranes, can form micelles (spheres). These can position themselves in an ordered arrangement that shows liquid crystal properties. As the concentration increases further, bilayers can form that can stack to form a lamellar (layered) phase liquid crystal (Figure A.24).

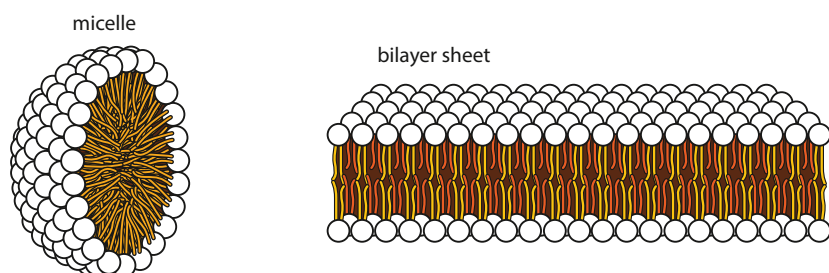


Figure A.24 A micelle and a bilayer that can lead to the liquid crystal state.

The **nematic liquid crystal phase** – molecules point, on average, in the same direction but are positioned randomly relative to each other (no positional order).

Molecules of substances that have liquid crystal properties are often rod-shaped and polar.

Thermotropic liquid crystals are pure substances but lyotropic liquid crystals are solutions.



Liquid crystal displays are everywhere in western society but the availability of modern technology is not uniform across the world. The 'One Laptop per Child (OLPC)' project is an attempt to make low-cost computers available to children in developing countries around the world.

Liquid crystal displays (LCD)

The ability of the molecules in a liquid crystal to transmit light depends on their orientation. Because the molecules are polar, their orientation can be controlled by applying a voltage. In the LCD, a weak electric field is applied to a thin film of the liquid crystal material held between two glass plates. By altering the orientation of the molecules using an electric field, the areas of the display that can and cannot transmit light – appearing light or dark – are controlled.

The ideal properties for a compound that can be used successfully in an LCD are:

- chemically stable
- exists in the liquid crystal phase over a suitable and wide range of temperatures
- polar so that they change orientation when an electric field is applied
- rapid switching speed between orientations.

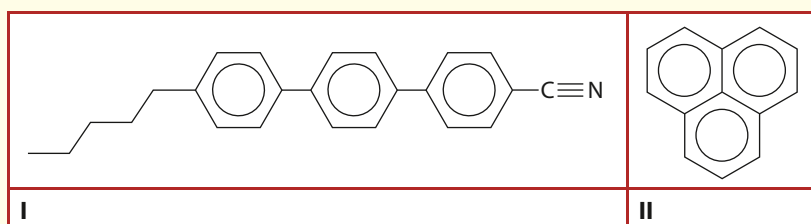
Current applications for LCDs include pocket calculators, digital watches and television and laptop computer screens. Liquid crystal displays are ideal for these purposes because of the requirement for only a tiny electric current – making them more energy efficient than other types of display. The main problems with liquid crystal displays are that they can be damaged fairly easily and they only operate over the temperature range in which the molecules exist in the liquid crystal phase – extreme hot and cold temperatures will temporarily disable an LCD.

Nature of science

Many great discoveries have been made by accident, rather than by making a hypothesis and testing it by experiment. Liquid crystals were accidentally discovered by Austrian botanist Friedrich Reinitzer when looking at derivatives of cholesterol – he realised that one of the molecules appeared to have two melting points. This was later found to correspond to a transition to or from the liquid crystal phase. Reinitzer's study of liquid crystals and their applications is also an example of scientists from different disciplines working together – Reinitzer later turned to physicist Otto Lehman for help in understanding the observed behaviour.

? Test yourself

- 9 Explain which of the following molecules would be more likely to show liquid crystal properties and be useful for a liquid crystal display.



A5 Polymers

Types of polymers

Polymers, commonly known as plastics, are formed when many small organic molecules (monomers) join together to form long-chain molecules (polymers). Polymers can be divided into two main classes – **thermoplastics** (thermosoftening polymers) and **thermosets** (thermosetting polymers).

Thermoplastics

Thermoplastics soften when they are heated and harden when they cool.

These polymers can be repeatedly heated and cooled, and remoulded into different shapes. The softening–hardening process is reversible. This is possible because thermoplastics consist of long-chain molecules with just intermolecular forces between the chains – intermolecular forces are overcome when thermoplastic polymers are heated (Figure A.25). These form again when the polymer cools. Examples of thermoplastics are polyethene and polychloroethene, PVC.

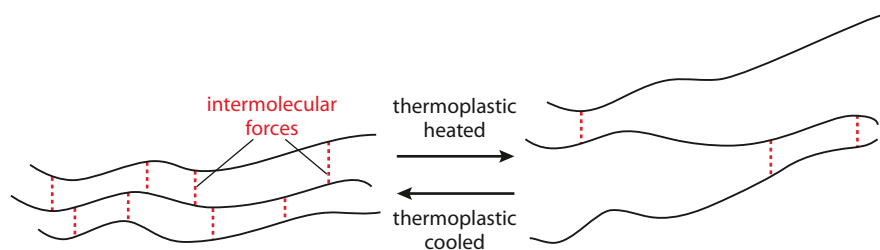


Figure A.25 Intermolecular forces are overcome when a thermoplastic is heated and form again when it is cooled.

Thermosets

The definition of a thermosetting polymer is:

a prepolymer in a soft solid or viscous state that changes irreversibly into a polymer network (thermoset) by curing.

To understand what this definition means, consider the formation of the thermoset when phenol reacts with methanal (Figure A.26).

This initial reaction of phenol with methanal in the presence of a catalyst produces various molecules that can join together into a prepolymer (a novolac) (Figure A.27).

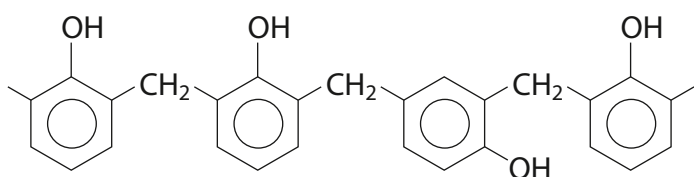


Figure A.27 Prepolymer for the formation of a phenol–methanal thermoset.

Learning objectives

- Understand the difference between thermoplastic and thermosetting polymers
- Understand what an elastomer is
- Understand the difference between the structures of HDPE and LDPE
- Understand how the properties of polymers depend on their structure
- Understand the difference between atactic and isotactic polymers
- Understand how the properties of polymers can be modified
- Use atom economy to evaluate the efficiency of a synthetic process

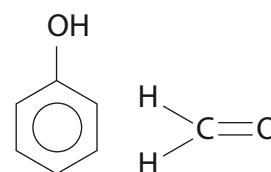


Figure A.26 Phenol and methanal.

This prepolymer is soft and has thermoplastic properties. When more methanal is added and the mixture heated, the prepolymer is cured to form a hard thermoset, which is made up of a three-dimensional cross-linked network (Figure A.28).

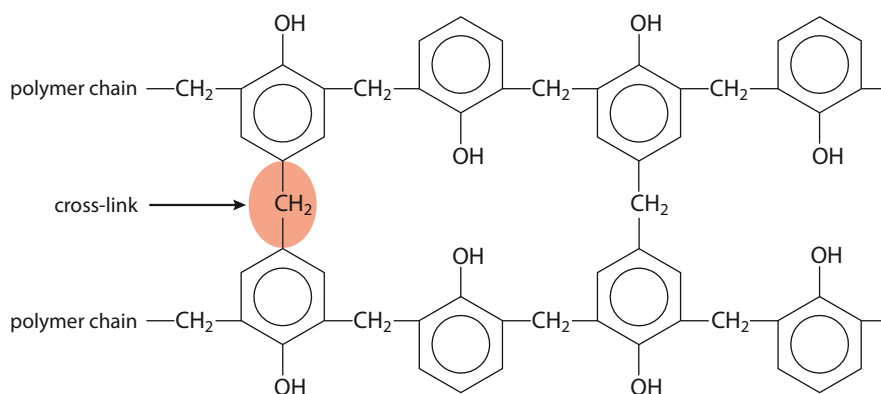


Figure A.28 Part of the structure of a phenol–methanal polymer.

This process is not reversible because covalent bonds (the cross-links) would have to be broken. So, once a thermoset is formed in a particular shape it cannot be moulded into any other shape. Moulding of the thermoset must therefore be done at the same time as curing.

Here we have seen curing as the result of heating (and the addition of a curing agent) but curing can also involve the action of electromagnetic radiation on a prepolymer. Examples of thermosets are Bakelite (a phenol–methanal polymer) and polyurethanes.

Elastomers

Elastomers are polymers that display rubber-like elasticity.

Elastomers, such as rubber, are flexible and can be stretched to many times their original dimensions by the application of a force. They will then return to (nearly) their original size and shape once the force is removed.

Elastomers are usually amorphous (non-crystalline) polymers (in the unstretched state) with some cross-linking between chains – so they are thermosets (there are also some thermoplastic elastomers). The polymer chains in elastomers are curled up (rather like a plate of spaghetti) and when a force is applied they tend to straighten. When the force is removed the chains return to their coiled arrangement.

Extension

The tendency of rubber to return to its original state can be understood in terms of entropy. When a piece of rubber is stretched, there is a decrease in entropy (the system becomes more ordered); when the stretching force is removed, the rubber returns to its original state, which has a higher entropy.

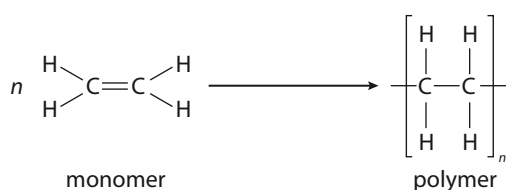
How the properties of thermoplastic polymers depend on structure

The properties of polymers depend on many factors – such as the length of the polymer chain (relative molecular mass), the degree of branching and the arrangement of groups on the polymer chain.

As the relative molecular mass of the polymer increases, it generally gets stronger and is able to be used at higher temperatures. This corresponds to stronger London forces between the chains as the relative molecular mass increases.

Polyethene

As discussed in Topic 10, ethene can be polymerised (addition polymerisation) to form polyethene:



Depending on the process used to make polyethene, different forms can be made.

Low-density polyethene

Low-density polyethene (LDPE) contains a high proportion of branching (Figure A.29a).

Highly branched polymer chains are less able to pack closely together, and therefore contact points between chains are reduced – this results in weaker London forces. Because the intermolecular forces between the chains are weaker, the polymer is more flexible and has lower tensile strength.

High-density polyethene

High-density polyethene (HDPE) has (virtually) no branching (Figure A.29b).

Lack of branching allows these polymer chains to pack together more tightly, increasing the density of the plastic. The more efficient packing of the chains increases the strength of London forces and so the chains are held together more tightly. This makes the polymer more rigid and increases its tensile strength. HDPE can also be used at a higher operating temperature. HDPE can be used for different purposes from those of LDPE because of its different mechanical properties.

Position of side groups

When propene is polymerised to polypropene, the repeating unit contains a methyl group side chain (Figure A.30).

Propene molecules can add together in different orientations so that the methyl (CH_3) groups on adjacent repeating units either on the same or

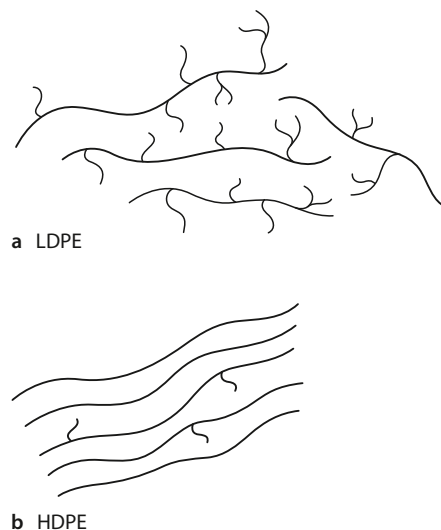


Figure A.29 The proportion of branching **a** in LDPE is significantly higher than in **b** HDPE and contributes to their different properties and uses.

Tensile strength refers to how well a material resists a stretching force without breaking.

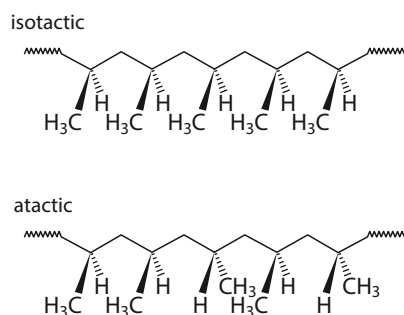


Figure A.31 Isotactic and atactic polypropylene.

Most commercially produced polypropylene is isotactic because of the carefully considered choice of catalyst during the polymerisation.

Atactic polypropylene is not manufactured commercially, and almost all atactic polypropylene is produced as a byproduct of isotactic polypropylene manufacture.

Syndiotactic polypropylene has the methyl groups on alternating sides of the polymer chain.

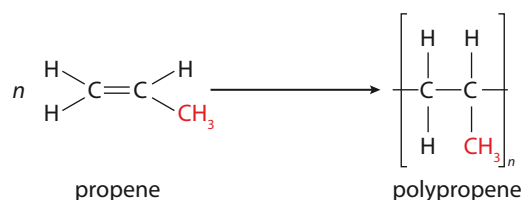


Figure A.30 Polymerisation of propene.

opposite sides of the polymer chain. The orientation of the methyl groups in relation to each other can affect the properties of the final material.

Isotactic polypropylene is the polymer of propene in which all the methyl groups are on the *same* side of the polymer chain. **Atactic polypropylene** contains methyl groups orientated in a *random* manner (Figure A.31).

The regular arrangement of methyl groups in isotactic polypropylene allows the chains to pack together more easily and, therefore, maximises the strength of London forces between chains. This is why isotactic polypropylene is crystalline, rigid and strong.

Isotactic polypropylene is used for a wide range of purposes including making flip-top lids, plastic kettles, crates, chairs and ropes.

The irregular positioning of the methyl groups in atactic polypropylene means that the polymer chains do not align themselves very well, and as a result the intermolecular forces between the chains are weaker. Atactic polypropylene is soft and rubbery, rather than rigid. It has a limited number of applications including its use as a roofing material, as a waterproof membrane and in paper lamination.

Plasticisers

Plasticisers are small molecules that are added to a polymer to increase its flexibility.

Unmodified PVC is a very rigid material used for guttering and piping – incorporation of phthalate plasticisers makes it flexible (Figure A.32). This flexibility (coupled with its impressive durability) makes plasticised PVC suitable for making garden hoses, flooring, inflatable structures and some clothing. The plasticiser molecules insert themselves between the polymer chains, forcing them apart and so reducing the strength of the intermolecular forces between them – this allows the chains to move more freely (Figure A.33).

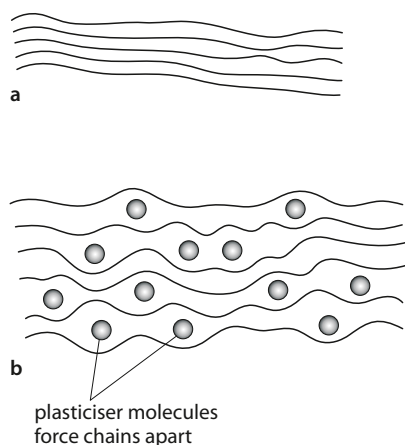


Figure A.33 PVC: **a** without plasticiser; **b** with plasticiser.

The most common plasticisers for PVC are **phthalates** (see Section A7).

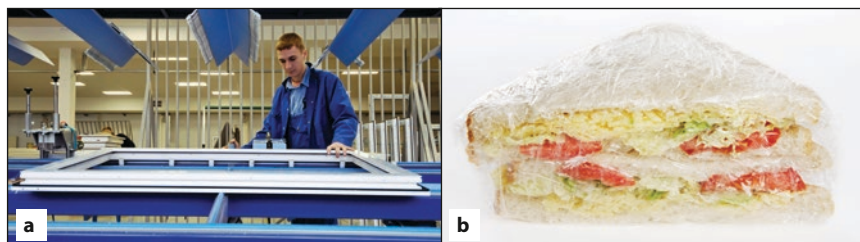


Figure A.32 **a** Unplasticised PVC is used to make window frames. **b** Plasticised PVC is used as a food wrap.

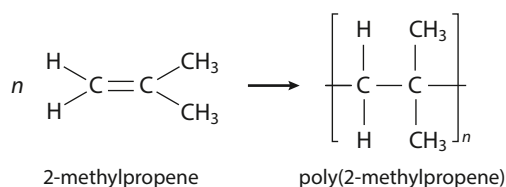
Expanded polymers

Polystyrene is a rigid, glassy plastic polymer in its unexpanded form. It is used to make plastic models, CD cases and disposable cutlery. However, most people would say that polystyrene is a white, low-density packaging material when asked to describe it.

This well-known form of polystyrene is the *expanded* form of the polymer – otherwise known as polystyrene foam or Styrofoam™. It is created by dissolving the volatile hydrocarbon pentane (C₅H₁₂) in the polymer during its initial manufacture. The beads of polystyrene that are formed during the polymerisation process are then heated in steam. The steam causes the pentane to vaporise and form gas bubbles within the polystyrene beads. This causes the beads to expand to about 60–70 times their original volume. Expanded polystyrene beads can then be pressed together in moulds to form sheets or appropriate shapes for packaging.

Polymerisation of 2-methylpropene

The polymerisation of 2-methylpropene can be represented by:



Normally the monomer molecules will add together in a head-to-tail fashion so that the methyl branches occur on every third carbon atom along the chain to give the polymer shown in Figure A.34a. However, occasionally a monomer unit can add the other way around to give variations in the chains, as shown in Figure A.34b.

There are no isotactic or atactic forms of poly(2-methylpropene) because there are two methyl groups on the same carbon and so different orientations on different sides of the chain are not possible (there are no chiral centres in the polymer chain).

Extension

Isotactic/atactic forms of a polymer chain require the presence of chiral centres in the chain.

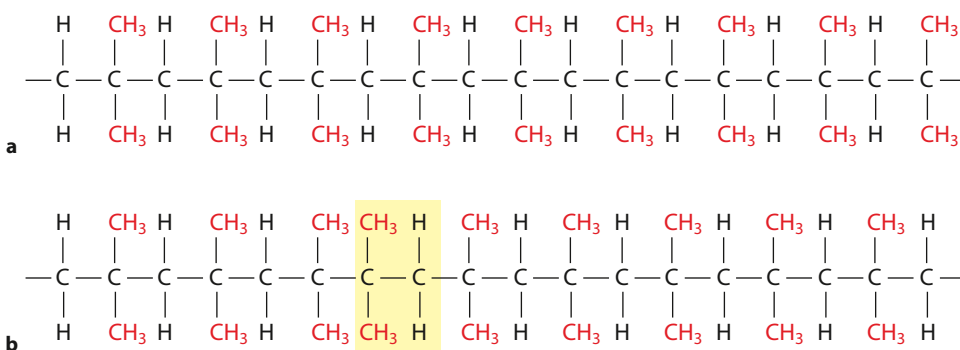


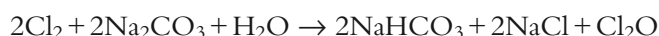
Figure A.34 **a** The polymer formed from 2-methylpropene where each monomer adds on to the chain in a regular head-to-tail arrangement. **b** The highlighted group added in a different orientation.

Atom economy

The idea of atom economy can be used as a measure of how efficient a particular reaction is in converting as much of the starting materials as possible into useful products. This is often used as a guide as to how 'green' a particular synthetic pathway is.

$$\text{atom economy} = \frac{\text{molar mass of desired products}}{\text{total molar mass of all reactants}} \times 100\%$$

We will use the preparation of Cl_2O :



to illustrate how atom economy is worked out.

The total molar mass of all reactants is calculated by working out the molar mass of each reactant, multiplying by the coefficients in the chemical equation and then adding them all up:

$$(2 \times 70.90) + (2 \times 105.99) + 18.02 = 371.80 \text{ g mol}^{-1}$$

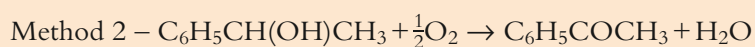
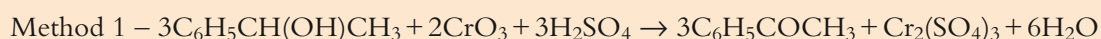
The desired product is Cl_2O , which has a molar mass of 86.90 g mol^{-1} .

$$\begin{aligned}\text{So, atom economy} &= \frac{86.90}{371.80} \times 100 \\ &= 23.37\%\end{aligned}$$

This is quite a low value because Cl_2O is only one of several products in this reaction – there is a lot of waste. Of course, if this were an industrial process, it would be greener if the other products were also used in some way and not simply wasted.

Worked example

A.2 Consider two different ways of making 1-phenylethanone from 1-phenylethanol:



Work out the atom efficiency for each process and suggest which is the more efficient.

Method 1

$$\begin{aligned}\text{Total molar mass of all reactants} &= (3 \times 122.18) + (2 \times 100.00) + (3 \times 98.09) \\ &= 860.81 \text{ g mol}^{-1}\end{aligned}$$

$$\begin{aligned}\text{Molar mass of desired product} &= 3 \times 120.16 \\ &= 360.48 \text{ g mol}^{-1}\end{aligned}$$

$$\begin{aligned}\text{Atom economy} &= \frac{360.48}{860.81} \times 100 \\ &= 41.88\%\end{aligned}$$

Method 2

$$\begin{aligned}\text{Total molar mass of all reactants} &= 122.18 + (0.5 \times 32.00) \\ &= 138.18 \text{ g mol}^{-1}\end{aligned}$$

$$\text{Molar mass of desired product} = 120.16 \text{ g mol}^{-1}$$

$$\begin{aligned}\text{Atom economy} &= \frac{120.16}{138.18} \times 100 \\ &= 86.96\%\end{aligned}$$

Method 2 has a much higher atom economy and is, therefore, much more efficient.

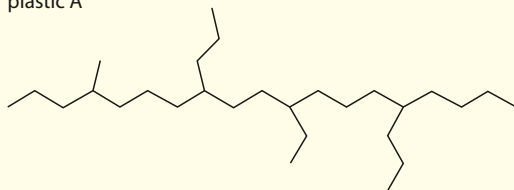
It is important to realise that atom economy is not the same as the *yield* of a reaction. Atom economy is a theoretical quantity based on a balanced equation and allows evaluation of how much waste product will be produced. The yield of a reaction is an experimental quantity worked out from how much of the desired product is actually made in a chemical reaction. In the calculation of atom economy above, it has been assumed that all reactions have 100% yield, which, in practice, will not be the case.

When evaluating how green or environmentally friendly a particular process is, both atom economy and yield must be considered – as well as several other factors such as how much energy must be supplied (usually as heat), the amount of solvents required, the nature of the solvents, disposal of solvents etc.

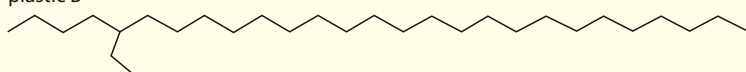
? Test yourself

- 10 Draw three repeating units of the addition polymer formed from but-2-ene and suggest whether it is possible to produce different forms.
- 11 The basic structure of the polymer chains in plastics A and B are shown below. Explain which will be denser and which will be more flexible.

plastic A



plastic B



- 12 Calculate the atom economy for each of the following reactions:
 - a $\text{CaC}_2 + \text{H}_2\text{O} \rightarrow \text{C}_2\text{H}_2 + \text{CaO}$, where the desired product is ethyne.
 - b $\text{C}_2\text{H}_4 + \text{PdCl}_2 + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{CHO} + \text{Pd} + 2\text{HCl}$, where the desired product is ethanal.
 - c $4\text{HgS} + \text{CaO} \rightarrow 4\text{Hg} + 3\text{CaS} + \text{CaSO}_4$, where the desired product is mercury.

Nature of science

Science involves an ever-changing body of knowledge and indeed Nobel prize winning physicist Richard Feynman comments on science as:

‘...the belief in the ignorance of experts.’

In the 1920s German chemist Hermann Staudinger, challenged the prevailing beliefs of many scientists by proposing the idea that substances such as rubber and cellulose are macromolecules made of many smaller units joined together by covalent bonds. He is widely regarded as the founder of polymer chemistry and although we take his ideas for granted now, his theories were not immediately accepted and were challenged by a lot of scientists.

As technology has advanced, the development of the use of techniques such as X-ray crystallography and scanning electron microscopy has enabled scientists to gain further understanding of the structures of polymers. The more scientists understand about the structure and properties of these substances, the better they are able to design new polymers to meet specific needs.

Learning objectives

- Explain what nanotechnology is
- Explain what is meant by molecular self-assembly
- Distinguish between physical and chemical methods of manipulating atoms
- Describe possible methods for synthesising carbon nanotubes
- Describe the structure and properties of carbon nanotubes
- Discuss some applications of nanotechnology
- Discuss the implications of nanotechnology

Polymerisation is not regarded as a self-assembly process because it involves the formation of covalent bonds and is not reversible.

A6 Nanotechnology

Nanotechnology is the production and application of structures, devices and systems at the nanometre scale. Generally, nanotechnology involves man-made particles or structures that have at least one dimension smaller than 100 nm.

The properties of materials change when their size falls below about 100 nm because of quantum effects and the fact that there is now a much higher ratio of atoms/molecules on the surface of the particle to those in the body of the material. Nanotechnology exploits these differences in properties.

We talk about ‘top-down’ and ‘bottom-up’ approaches to nanotechnology.

- In the top-down approach, etching and machining are used to create a nanoscale structure by making things smaller – computer chips are created by a top-down approach.
- In the bottom-up approach, atoms or molecules are manipulated by either chemical or physical means to create nano-sized structures – starting with the smallest possible particles we build up a larger structure.

Molecular self-assembly is a bottom-up approach to producing nanoparticles, where molecules come together reversibly and spontaneously to create a larger structure. This may occur when molecules attach themselves to a surface or when particles come together spontaneously in solution. An example of molecular self-assembly is when soap molecules in solution come together to form a micelle or a bilayer (page 21). Molecular self-assembly does not include building up larger molecules using chemical reactions that involve the formation of covalent bonds, but rather how molecules come together in a specific way due to intermolecular forces such as London forces, hydrogen bonds and electrostatic interactions. Two strands of DNA coming together to form a double helix is an example of molecular self-assembly, whereas a protein folding into a specific three-dimensional structure is an example of intramolecular self-assembly.

Chemical and physical methods of creating nanoscale structures

Atoms can be manipulated or moved into position by both chemical and physical techniques. Chemical manipulation relies on the use of specific chemical reactions or interactions to position the atoms in a molecule. There are many examples of the formation of nanoscale structures using chemical reactions – for example, amino acids can be joined together into polymer chains by the formation of covalent bonds between the individual units, which will then fold into specific conformations under the influence of hydrogen bonding or other intermolecular forces. Using more complex chemistry, rotaxanes (Figure A.35) can be made in which a cyclic molecule is held on a rod-shaped molecule without covalent bonding. A rod-shaped molecule has bulky stopper groups at each end so the ring cannot come off and these nanostructures have potential uses as molecular machines.

Physical manipulation can be used to position atoms in particular places. Using a **scanning tunneling microscope**, it has been possible to pick up individual atoms and move them to different places on a surface. This is how the logo in Figure A.36 was created by moving atoms around on a surface.

Physical techniques: atoms are manipulated and positioned to specific requirements.

Chemical techniques: atoms are positioned in molecules using chemical reactions.



In Figure A.36 are we actually ‘seeing’ atoms? Does a photograph like that shown in Figure A.36 mean that we know, without any doubt, that atoms exist?

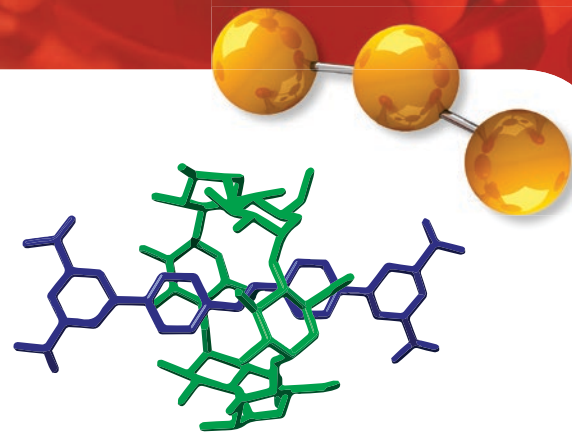


Figure A.35 The structure of a rotaxane.

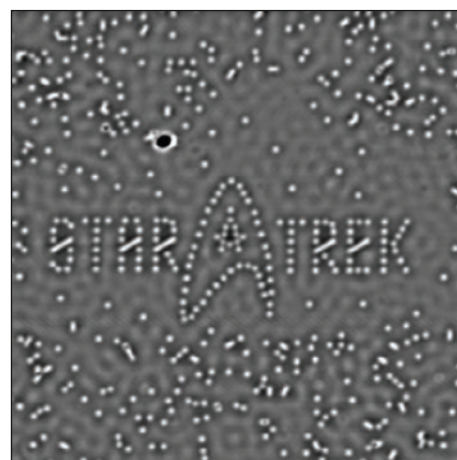


Figure A.36 The Star Trek logo was created by IBM by manipulating individual atoms.

Carbon nanotubes

Carbon nanotubes are **allotropes** of carbon and have a structure that is analogous to a single layer of graphite (graphene) rolled into a tube to create a cylinder of hexagons of carbon atoms (Figure A.37). It is possible to create single-walled nanotubes (SWNTs) and multi-walled nanotubes (MWNTs). In MWNTs, there is a concentric arrangement of two or more nanotubes. The diameter of a single-walled carbon nanotube is typically about 1–2 nm and lengths of up to about 20 cm have been reported (but most are much shorter). Some carbon nanotubes are closed at the end (capped) and some are open. In order to allow the ends of the tubes to be sealed, pentagons must also be present in the structure.

Synthesis of carbon nanotubes

There are several different methods for producing carbon nanotubes.

Arc discharge

This uses two carbon rods placed very close together (1–2 mm) often in a low-pressure inert-gas (e.g. helium, argon or nitrogen) atmosphere – although synthesis can also take place in the open air. A high current (typically around 100 A) creates an arc between the electrodes. The high

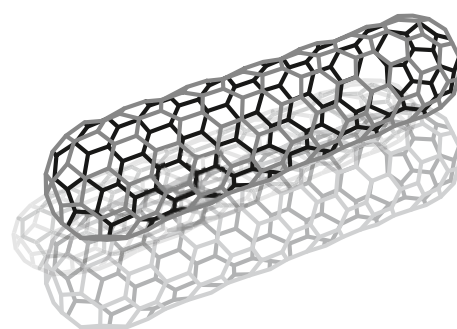


Figure A.37 A capped single-walled carbon nanotube. Note the hexagons that make up the main body of the tube and the inclusion of pentagons at the capped ends.

temperature of the arc causes one electrode (anode) to vaporise and the carbon is then deposited as carbon nanotubes on the other electrode (cathode). A catalyst is sometimes used in the process.

A variation on the arc discharge method is to generate an arc between two metal electrodes (e.g. nickel) in a liquid hydrocarbon solvent (e.g. methylbenzene). The hydrocarbon is decomposed in the arc and rod-shaped deposits are formed on the anode. The average oxidation number of carbon in methylbenzene ($\text{C}_6\text{H}_5\text{CH}_3$) is -1.14 , but because the carbon nanotubes are a form of elemental carbon, the oxidation number of carbon in them will be zero. Therefore this process involves the oxidation (increase in oxidation number) of carbon (from the hydrocarbon) at the anode.

Chemical vapour deposition (CVD)

A carbon-containing gas (e.g. methane, ethyne) is heated to a high temperature (usually above 500°C but low-temperature synthesis is also possible). This is done in the presence of a metal nanoparticle catalyst (typically iron, cobalt or nickel) supported on a substrate (e.g. a zeolite). At the high temperatures involved, the covalent bonds in the carbon-containing compound are broken at the surface of the catalyst and nanotubes can be built up. Oxygen should be excluded from the reaction mixture because the oxygen will react with the carbon/organic compound (to form CO or CO_2).

High-pressure carbon monoxide deposition (HIPCO)

A mixture of carbon monoxide and iron pentacarbonyl, $\text{Fe}(\text{CO})_5$, is fed into a reaction vessel at high pressure ($0.3\text{--}0.5\text{ MPa}$) and high temperature ($900\text{--}1100^\circ\text{C}$). At these temperatures the $\text{Fe}(\text{CO})_5$ decomposes into iron and carbon monoxide. The iron atoms come together to form clusters on which the carbon nanotubes form by a disproportionation (the same species oxidised and reduced) reaction:



The iron clusters/nanoparticles act as the catalyst for the formation of the carbon nanotubes.

Properties of carbon nanotubes

A single nanotube is an extremely strong structure because only covalent bonds are present between the atoms. To break apart a nanotube, covalent bonds must be broken and these are very strong. Nanotubes are among the strongest materials ever created – they are much stronger than steel and also have a much lower density.

It may be possible to make bundles of these tubes that also show exceptional mechanical properties, but the challenge is to produce sufficiently long and well-aligned fibres so that the strength of the bundle as a whole relies more on the strength of the covalent bonds between atoms in the individual nanotubes rather than the forces between tubes. This can be compared with graphite – although each individual layer (graphene) has exceptional mechanical properties (it is very strong), when the layers come together to form graphite, a soft substance (used as a dry lubricant) is produced because of the weak forces of attraction between the layers.

As in graphite and graphene, the carbon atoms in nanotubes form only three bonds, so there is one electron that is not involved in bonding present on each carbon atom. These electrons become delocalised over the whole structure and so carbon nanotubes are able to conduct electricity.



The degree to which a carbon nanotube conducts electricity depends on its length, as a result of changes in the behaviour of the delocalised electrons at the nanoscale – quantum-level effects predominate at the nanoscale. Some nanotubes are full conductors, whereas others have semiconducting properties. The use of nanotubes in electronic circuitry has long been proposed.

Applications of nanotechnology

Carbon nanotubes have been used in making composite materials to produce strong, low-density materials that can be used to make, for example, parts of bicycles and other sporting equipment and body armour. Carbon nanotubes have potential to be highly effective heterogeneous catalysts – they have a large surface area because both the outside and inside surfaces are available for binding to reactants. Nanotubes could be developed into specialist filters in which the diameter of the tubes is set to allow the passage of particles up to a certain maximum size – for example, in the desalination of water by allowing small water molecules through but excluding larger chloride ions.

Worries about nanotechnology

There are some serious concerns about nanotechnology in the health arena. Determining the toxicity of nanotubes and nanoscale particles is difficult because their properties depend on their size. There is speculation that they could cross cell membranes and thereby induce harmful effects. Concerns have surfaced that the human immune system would not be able to recognise particles on the nanoscale and would be defenceless against them. The similarity between carbon nanotubes and asbestos threads has been noted and has led to worries that nanotubes might be able to cause respiratory and other health issues in the way that asbestos does.

The technology is so new that not enough is known about the potential implications for human health. New materials being created may have new and unforeseen health risks – thorough testing and regulation will be essential. Government regulatory bodies, and the industry itself, will need to take responsibility for the safe introduction of nanotechnology around the world.

Nature of science

There have been many cases in the past where advances in science and technology have inadvertently caused major environmental and health problems. Scientists have then worked to try to solve/reduce the effects of these problems. We are, however, now much more aware of the potential problems that can arise and scientists have a moral responsibility with nanotechnology to consider the possible consequences of their work as they are doing it rather than afterwards, when it is often too late. Political and economic factors can, however, often drive decisions so it is not the case that scientists are always totally objective when drawing conclusions.

The development of new apparatus and technology has been fundamental in the growth of nanotechnology as an exciting new branch of science. Without sophisticated equipment such as the scanning tunnelling microscope, which has allowed the manipulation of atoms, nanotechnology probably would not exist.

Carbon nanotubes also have many other interesting properties such as very good thermal conductivity along the tube, an ability to absorb certain frequencies of electromagnetic radiation etc.

Other proposed applications of nanotechnology include hydrogen storage in hydrogen-powered vehicles, synthesis of chemical nanowires, visual displays, solar cells, stealth technology and many, many more.

Learning objectives

- Understand that most plastics are non-biodegradable
- Derive chemical equations for the combustion of plastics
- Understand that toxic products are released when PVC is burned
- Describe the structure of dioxins
- Understand some of the health risks associated with dioxins
- Compare the structures of dioxins and PCBs
- Understand the health effects of using plasticisers in polymer production
- Discuss the environmental impact of the use of plastics
- Understand that recycling plastics is more labour-intensive and difficult than for many other materials
- Understand that plastics must be sorted according to type before recycling can be done
- Understand how resin identification codes for some plastics can be identified from IR spectroscopy

This equation is identical to the one for combustion of the monomer.

Polychloroethene can also be called polyvinyl chloride (PVC)

A7 Environmental impact – plastics

Plastics derived from alkenes, such as polyethene, polypropene, polychloroethene, are non-biodegradable, which means that they cannot be broken down by microorganisms when, for instance, they are buried in soil. They are non-biodegradable because of the strong carbon–carbon covalent bonds in the polymer chain.

Because plastics are non-biodegradable they are difficult to dispose of – the three main methods for dealing with waste plastic are burying in a landfill site, incineration and recycling.



Certain European countries, such as Denmark and Switzerland, incinerate large proportions of their waste but other countries, such as Spain, Finland and Ireland, predominantly use landfill sites.

Combustion of plastics

The products formed by the combustion of plastics depend on various factors such as:

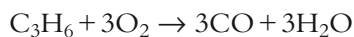
- the composition of the plastic
- the availability of oxygen
- the temperature.

In a good supply of oxygen, hydrocarbon plastics should undergo complete combustion to form carbon dioxide and water. For example, the combustion of polyethene in a good supply of oxygen could be written as:

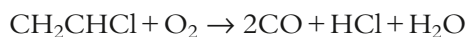


This equation ignores any end groups on the polymer chains.

In the presence of a limited supply of oxygen carbon monoxide (toxic) and soot (carbon) can be formed. We can represent the combustion by an equation based on the repeating unit of a polymer. Therefore the incomplete equation of polypropene to produce carbon monoxide can be represented as:



When polymers containing chlorine are burned, hydrogen chloride and other products can be formed. A simplified equation for the combustion of polychloroethene is shown here:

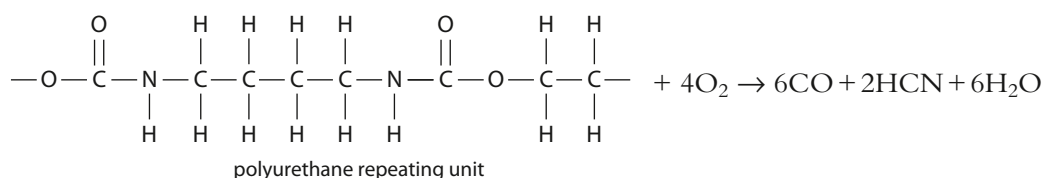


There is not enough chlorine in this polymer for all the hydrogen to be converted to HCl. However, if polydichloroethene is burned, this would be theoretically possible:



Polymers containing fluorine will produce hydrogen fluoride etc.

When polymers containing nitrogen are burned, hydrogen cyanide (HCN) and nitrogen oxides (NO_x) can be formed. The following equation represents the burning of a polyurethane polymer and the formation of hydrogen cyanide.



Nitrogen oxides may be formed by the oxidation of the nitrogen in the polymer or, if the temperature is high enough, in the reaction between nitrogen and oxygen in the air. NO is formed first, which can be further oxidised to NO₂.

The above equations illustrate the formation of some simple products of combustion of polymers – but in reality combustion produces a complex mixture of compounds. For instance, burning of polychloroethene can also produce dioxins (see later).

Many toxic substances may be formed in house fires – a burning PVC shower curtain can release HCl and dioxins; polyurethane foams used in furniture can release hydrogen cyanide and isocyanates, which are both potentially fatal.

PVC is often used as the insulating material for electrical cables. Where there is potentially a significant danger to the public because of cable fire, such as in aeroplanes and in public buildings, PVC can be replaced with low-smoke, zero-halogen cabling, which gives off very little smoke when it is burned and does not produce toxic halogen-containing compounds.

? Test yourself

- 13 Suggest the products of burning the polymers formed from each of the following alkenes in a good supply of oxygen:
- $\text{H}_2\text{C}=\text{C}(\text{CH}_3)\text{COOCH}_3$
 - $\text{H}_2\text{C}=\text{CHF}$
 - $\text{H}_2\text{C}=\text{CHCN}$

Dioxins

The simplest dioxin structures are shown in Figure A.38. They consist of a six-membered heterocyclic ring with two oxygen atoms – 1,4-dioxin is the more common form.

The term ‘dioxin’ is, however, usually used to describe the polychlorinated derivatives of the compound in Figure A.39. These are also called polychlorinated dibenzodioxins (PCDDs).

Dioxins are produced as byproducts in the manufacture of some chlorinated organic compounds. They are also formed if the temperature is not high enough (below about 1200 °C) when waste materials containing organochlorine compounds are incinerated. The most toxic of these derivatives is called 2,3,7,8-TCDD (or just TCDD or 2,3,7,8-tetrachlorodibenzo-1,4-dioxin, or just dioxin). The structure of this is shown in Figure A.40.

Dioxins are chemically unreactive and do not decompose in the environment. They accumulate in the fatty tissue of animals and are passed up the food chain. The main exposure of humans to dioxins comes from food – meat, fish and dairy products. Dioxins act by disrupting the correct action of hormones, which can affect growth and the functions

A heterocyclic ring is one that contains atoms other than carbon incorporated into a ring of carbon atoms.

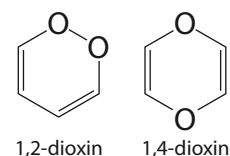


Figure A.38 The basic dioxin structure.

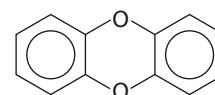


Figure A.39 The structure of dibenzo-1,4-dioxin.

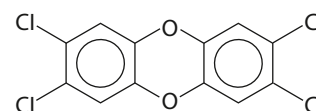


Figure A.40 Dioxin.

Exam tip

The structures of dioxins, PCBs etc. may be shown with benzene rings with a delocalised ring (circle in the middle) or with alternating single and double bonds.

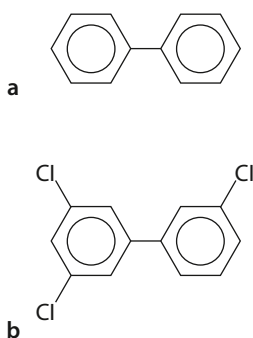


Figure A.41 **a** The basic biphenyl structure; **b** a PCB.

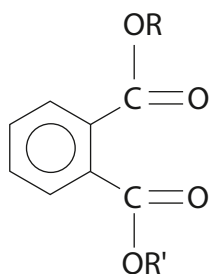


Figure A.42 The basic structure of a phthalate ester – they are esters of benzene-1,2-dicarboxylic acid.

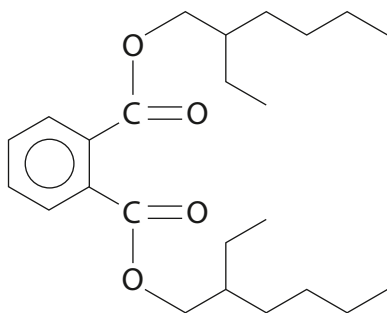


Figure A.43 DEHP.

of many systems in the body. The effects of exposure to dioxins are still being studied but adverse health effects include liver damage and a skin disease called chloracne. Studies on animals have shown that dioxins can cause cancer and they are classified as human carcinogens. However, the World Health Organisation states that 'TCDD does not affect genetic material and there is a level of exposure below which cancer risk would be negligible.'

PCBs

PCBs are **polychlorinated biphenyl** compounds. There are 209 possible PCBs where between one and ten hydrogen atoms in the basic biphenyl structure (Figure A.41a) are replaced by chlorine atoms. An example of a PCB is shown in Figure A.41b.

PCBs are similar to the polychlorinated dioxins discussed earlier in that their molecules contain two benzene rings (phenyl groups) and chlorine atoms – but PCBs do not contain oxygen or a heterocyclic ring.

PCBs are chemically inert, non-flammable and stable at high temperatures. They were used in making electrical transformers and capacitors because of their high electrical resistance – so factories making these would have discharged PCBs into the environment. PCBs have not been manufactured in the USA since 1979 – however, because they are unreactive they persist in the environment for a long time. They also accumulate in fatty tissue and have been linked to low reproduction rates among some marine animals and are thought to be carcinogenic in humans. PCBs can be passed from mother to child in milk.

Phthalate esters

Phthalate esters (Figure A.42) are used as plasticisers in the polymer industry – for example in the manufacture of PVC (polychloroethene). These molecules sit between the polymer chains and increase the flexibility of the polymer. Because they are not covalently bonded to the polymer chains they can be released into the environment when the plastic is used. There are health concerns about the use of these compounds.

One of the most common phthalate esters used as a plasticiser is DEHP (bis(2-ethylhexyl) phthalate or di(2-ethylhexyl) phthalate) (Figure A.43).

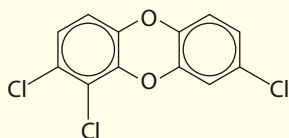
DEHP may be present in many household articles such as packaging material, food wrap, furniture upholstery, floor coverings, children's toys and shower curtains. Particular concern has arisen about its use in food packaging when the food involved has a high fat content because DEHP is fat-soluble. DEHP is also used in a variety of medical products such as IV tubes and blood bags.

The use of DEHP is controlled in many countries. Its health effects are not clear – it can cause cancer in mice and rats but the situation with humans is not conclusive and it is classified differently by different organisations. It is regarded in some countries as a substance that could probably cause cancer but the International Agency for Research on Cancer believes that there is not enough evidence to make a decision either way about its carcinogenicity. Phthalates such as DEHP have been associated with disruption of the endocrine system in humans and are believed to have an adverse effect on sexual development.

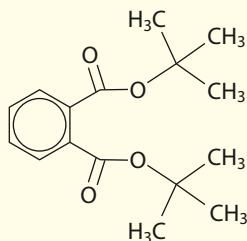
? Test yourself

14 Classify each of the following as a PCB, PCDD or phthalate ester.

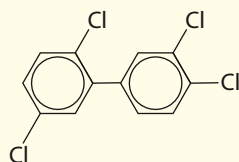
a



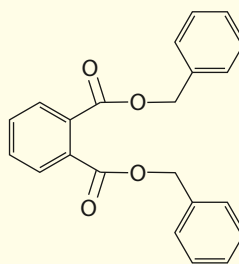
b



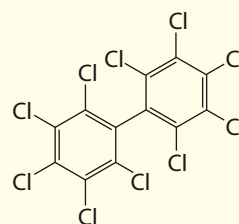
c



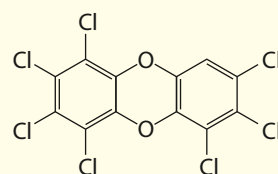
d



e



f



Are there limits on the pursuit of knowledge? To what extent should environmental, ethical issues, etc. determine which areas of knowledge are pursued?

The environmental impact of the use of plastics

There are many factors that have to be considered when discussing the environmental impact of the use of plastics. It is far from a simple debate and we cannot simply say that plastics are bad for the environment. Some relevant points are summarised below.

- Plastics are made from crude oil, which is a finite resource.
- Plastics persist for a long time in the environment.
- Plastics are usually non-biodegradable, so they have to be disposed of by burying in landfill sites or by incineration. Landfill sites can be unsightly, smelly and noisy and take up large areas of land. Incineration is more expensive and can create toxic chemicals – incineration also results in the production of carbon dioxide, a greenhouse gas.
- Many plastics containers need less energy to make than an equivalent one in glass or aluminium.
- The use of plastics in making insulation materials reduces energy losses.

- The energy liberated in the incineration of plastics has useful purposes – heating buildings for example.
- Plastic containers are light and less energy is needed to transport them – for example, fizzy drinks in plastic bottles rather than in glass bottles.
- Plastic packaging can reduce food wastage – if less food is wasted then less energy needs to be used to produce more.
- Plastic pipes for water, gas, sewage and communication cables do not rust and therefore do not have to be replaced as often – this reduces energy consumption.
- Plastic debris in the environment can cause harm to birds and marine animals.

Some of the health effects of the use of plastics have been discussed above.

Recycling plastics

An alternative to burying plastics or incinerating them is to recycle them. Recycling of plastics is more labour intensive and difficult than for many other materials. There are many different types of plastics and before recycling they must first be separated from each other. The plastic is then shredded and washed, and then melted, extruded through holes and chopped into pellets. These pellets can then be remelted and moulded to make new products.

Not all plastics are equally easy to recycle. The most commonly recycled plastics are PET and HDPE because these contain the smallest amount of additives. Some plastics (e.g. PVC), containing higher proportions of additives, may require more energy to purify than would be required to make them from crude oil. Thermosets have cross-linking between polymer chains, which means they cannot be remelted and reformed – they are often crushed and used as insulation.

Pyrolysis (cracking) may also be used in recycling plastics. Pyrolysis involves heating plastics in the absence of oxygen to split them up into smaller molecules that can be used as a chemical feedstock to make new plastics or as a fuel. Thermosets can also be processed in this way.

PET, the main plastic used in fizzy-drink bottles, can either be remelted and formed into new bottles or can be hydrolysed to break it down into its monomers.

Recycling plastic is expensive – more expensive than dumping it in landfill sites – but it can reduce energy consumption, emissions of carbon dioxide, the need for more landfill sites and also conserve crude oil, which is an extremely valuable natural resource.

Sorting plastics

It is very important that plastics are sorted completely into their different types before recycling. Any sort of contamination can lower the quality of the plastic produced – or even ruin a whole batch. Plastics are usually sorted by hand in a very labour-intensive process. Sorting is aided by resin identification codes (RIC) on the plastic objects – see Table A.3.

Each different type of plastic is processed separately.








Resin identification code (RIC)	Plastic type	Resin identification code (RIC)	Plastic type
 PETE	polyethylene terephthalate	 PP	polypropylene/polypropene
 HDPE	high-density polyethylene/ polyethene	 PS	polystyrene/poly(phenylethene)
 PVC	polyvinyl chloride/ polychloroethene	 OTHER	other
 LDPE	low-density polyethylene/ polyethene		

Table A.3 Resin identification codes – PETE may also be labelled 'PET'.

Using IR spectroscopy to identify different plastics

Infrared spectroscopy can be used to identify different types of plastic. Some polymers contain characteristic functional groups and this aids identification. For instance, the structure of PET (polyethylene terephthalate) is shown in Figure A.44.

This polymer contains an ester group and therefore we would expect to see a band in its infrared spectrum in the range $1700\text{--}1750\text{ cm}^{-1}$ due to C=O stretching and a band in the region $1050\text{--}1410\text{ cm}^{-1}$ due to the C–O stretching, although the latter is much harder to spot. If the infrared spectra of polyethene and PET are compared (Figure A.45) the difference can be seen clearly.

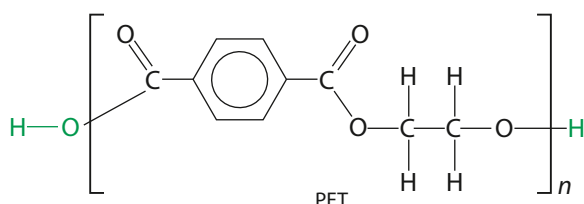


Figure A.44 The structure of PET showing the repeating unit.

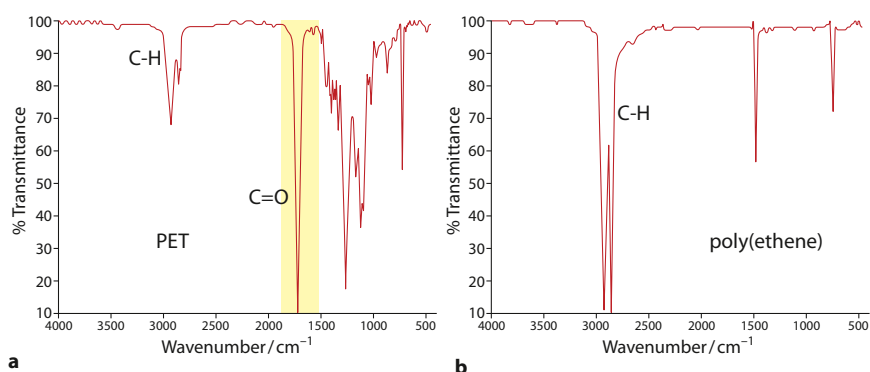


Figure A.45 IR spectra of **a** PET; **b** polyethene.

It would be much more difficult to distinguish between polypropene and polyethene using this method because they both contain the same bonds (C–C and C–H). It can also be difficult to identify a plastic such as polychloroethene because the presence of a band in the $600\text{--}800\text{ cm}^{-1}$ region is no guarantee of a C–Cl bond – there are many other vibrations that give rise to absorptions in this region.

The infrared spectrum of polystyrene is shown in Figure A.46. The absorption bands at 1600 cm^{-1} and just below 1500 cm^{-1} are due to the vibrations of C=C in the benzene ring – these bands are very characteristic of benzene rings.

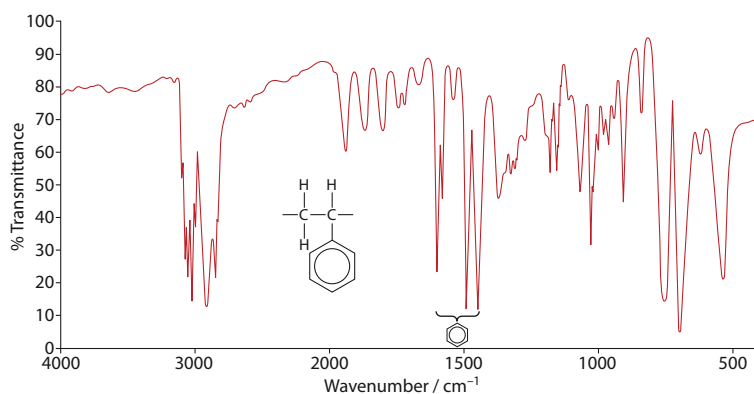


Figure A.46 The infrared spectrum of polystyrene ([poly(phenylethene)]) and the repeating unit.

Nature of science

There must be some evidence to be able to draw scientific conclusions. There are lots of rumours about the health effects of substances such as DEHP, but they are not backed up by clear evidence. Carrying out studies involving human health issues can be difficult because scientists must act ethically and are governed by rules and regulations about what they can and cannot do. It is difficult to carry out tests with large groups under controlled conditions and scientists must try to draw conclusions by the statistical analysis of data. This can lead to contradictory findings in different studies.

The development and use of polymers/plastics has grown enormously over the last 100 years but it is only now that we are beginning to realise some of the risks involved both in terms of the environment and human health. In the past scientific research and development seems to have continued irrespective of these issues but nowadays scientists are much more aware of the impact that new materials can have on health and environment – the risks and benefits of new products are considered both by the scientists themselves and governmental organisations.

? Test yourself

15 Match the absorptions in infrared spectra given below to possible resin identification codes:

- a around 1600 cm^{-1}
- b $1700\text{--}1750\text{ cm}^{-1}$
- c $2850\text{--}3090\text{ cm}^{-1}$
- d $600\text{--}800\text{ cm}^{-1}$

Exam tip

PET/PETE is the only polymer, whose RIC is given specifically, that will have an absorption in the range $1700\text{--}1750\text{ cm}^{-1}$. However, polymers in the 'others' category could also have this absorption.

A8 Superconducting metals and X-ray crystallography (HL)

Superconductors are materials that have zero electrical resistance below a **critical temperature**.

Superconductivity can be observed in some metals. The resistance of all metals decreases with temperature but for some metals, such as tin and aluminium, the resistance drops to zero below a critical temperature (Figure A.47). The critical temperature of tin is about 4 K. Not all metals show superconductivity and some metals still have some resistance at very close to absolute zero.

Metals conduct electricity because the delocalised electrons are able to move through the structure. Resistance in metals arises because these electrons collide with the positive ions in the lattice (Figure A.48). As temperature decreases, the metal ions vibrate less and, therefore, there is essentially a smaller cross-section for the electrons to collide with and the resistance decreases.

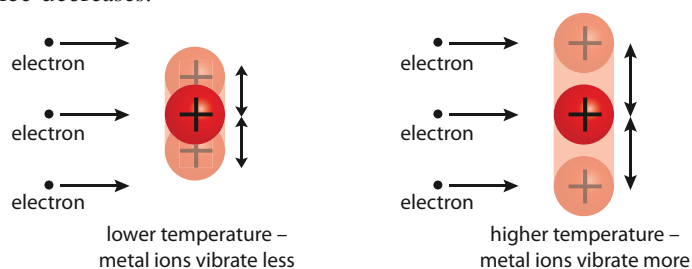


Figure A.48 Metal ions vibrate less at lower temperatures and the electrons are better able to pass through the structure.

Many alloys and some ceramics can also exhibit superconductivity. Some of these, such as $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ ($0 < x \leq 0.6$), can superconduct up to about 93 K depending on the proportion of oxygen. Substances such as this are called high-temperature superconductors.

The Bardeen–Cooper–Schrieffer (BCS) theory of superconduction

This theory explains superconductivity in terms of the interactions between a material's delocalised electrons and its lattice that allow electrons to form pairs that can move unhindered through the material. At a simple level, when an electron passes through a lattice at a low temperature, it attracts the positive ions around it slightly. This creates a region in the lattice with slightly more positive charge, which attracts another electron. The two electrons interact via the lattice and are weakly bound together to form a **Cooper pair**. It is the formation of these Cooper pairs that causes superconductivity – these electron pairs move freely through the lattice structure. The electrons are not very close together and the attraction via the lattice is stronger than the repulsion between the electrons due to their like charges. If the temperature increases, the electrons cannot cause distortion of the lattice in the same way because the lattice vibrations become too strong, Cooper pairs cannot be formed and the superconductivity disappears.

Learning objectives

- Understand what is meant by a superconductor
- Understand how electrical resistance in metal arises
- Explain superconductivity in terms of the Bardeen–Cooper–Schrieffer theory
- Understand the Meissner effect
- Understand the difference between Type 1 and Type 2 superconductors
- Understand that crystal lattices can be described in terms of a unit cell that repeats throughout the structure
- Understand how to work out the number of atoms in a unit cell
- Understand the term coordination number
- Understand that X-ray diffraction can be used to work out the structure of metallic and ionic lattices
- Apply the Bragg equation to simple cubic structures
- Calculate the density of a metal from atomic radii

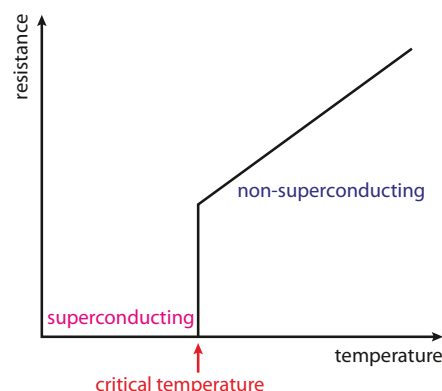


Figure A.47 The variation of resistance with temperature for a metal such as tin.

Magnetic properties of superconductors

Superconductors expel external magnetic fields. When a superconductor is below its critical temperature and it is exposed to a magnetic field, the movement of electrons on the surface of the material creates a magnetic field that exactly opposes the external field and prevents the penetration of the field into the material (Figure A.49). This is known as the Meissner effect.

The Meissner effect is the reason why superconducting materials can levitate in a magnetic field or cause magnets to levitate.

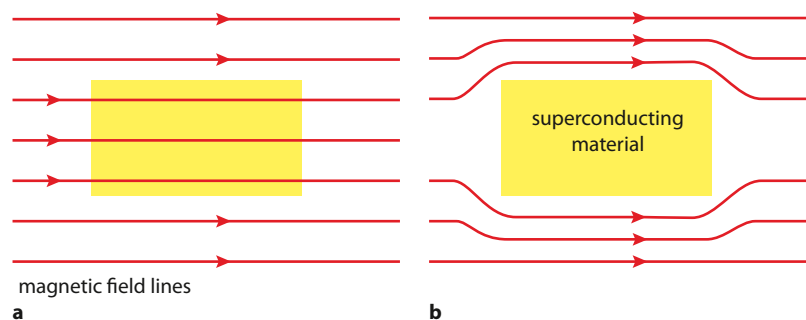


Figure A.49 **a** A superconductor above its critical temperature in a magnetic field; **b** when the superconductor is cooled to below its critical temperature the magnetic field is expelled.

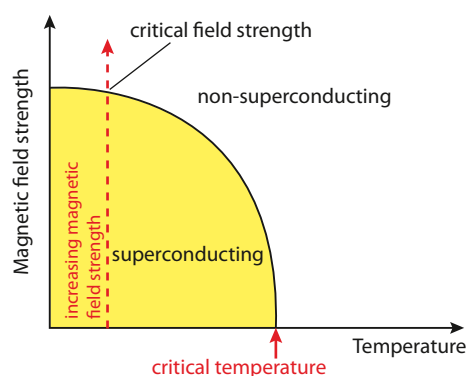


Figure A.50 Magnetic field strength against temperature for a Type 1 superconductor. As magnetic field strength is increased at a temperature below the critical temperature there is a sharp transition from superconducting to non-superconducting.

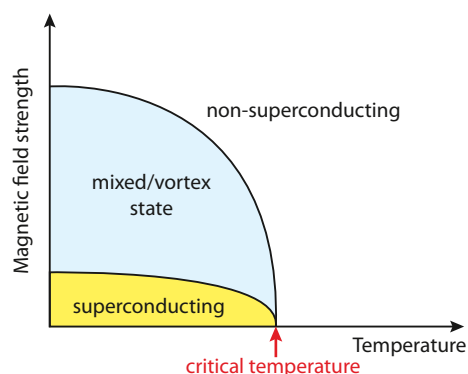


Figure A.51 Magnetic field strength against temperature for a Type 2 superconductor.

Type 1 and Type 2 superconductors

Superconductors can be divided into two classes depending on how superconductivity is affected by a magnetic field. Superconductivity can be destroyed by the application of a strong enough magnetic field. Type 1 superconductors have a sharp transition from superconducting to non-superconducting as the magnetic field strength is increased at a particular temperature (Figure A.50).

The magnetic field strength that is required to destroy the superconductivity is called the critical field strength (H_c) and depends on the material and the temperature. Most pure metallic element superconductors, such as tin, are Type 1 superconductors.

Type 2 superconductors, which include alloys and all the high-temperature ceramic superconductors, have two critical field strengths at temperatures below the critical temperature and do not show a sharp transition from superconductivity to non-superconductivity (Figure A.51). In the yellow region the material exhibits superconductivity, but as the magnetic field strength is increased at a particular temperature there is a transition to a mixed state (blue region) where there is some penetration of the magnetic field into the substance and there are superconducting and non-superconducting regions. As the magnetic field strength is increased further, a higher critical field strength is reached, above which there is no superconductivity. The transition from superconducting to non-superconducting as the magnetic field strength is increased can therefore be described as gradual.

The difference between Type 1 and Type 2 superconductors arises from their behaviour in a magnetic field.

Lattice structures

Metals, ionic compounds and covalent compounds can form crystal structures. A crystal contains a regular repeating array of atoms, ions or molecules. Here we will consider some of the structures adopted by metals.

The particles in crystals can be arranged in many different ways and arrangements can be described by using the idea of a unit cell.

Consider the two-dimensional lattice structure in Figure A.52. The unit cell can be translated horizontally or vertically each time by one unit cell length to build up the whole pattern.

If we extend this to three dimensions by putting identical layers directly above this layer, we get a cubic lattice. Now, the unit cell is a cube and if it is translated throughout the structure the whole lattice can be built up (Figure A.53).

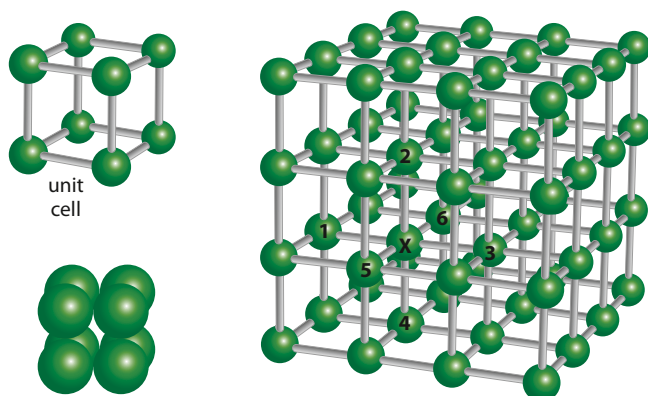


Figure A.53 Two different representations of a simple cubic unit cell and the cubic lattice.

Each atom in the cubic lattice has six nearest neighbours (look at the atom marked with 'X' in Figure A.53) – it is said to have a coordination number of six.

Unit cells are usually shown with an atom (or ion) at each corner but, if you look at the simple two-dimensional square unit cell, it can be seen that only $\frac{1}{4}$ of each atom is in the unit cell (Figure A.52). In the three-dimensional cubic lattice in Figure A.53 it can be seen that eight cubes come together at any point, and therefore there is only $\frac{1}{8}$ of each atom at the corner of the cube in the unit cell. The simple cubic unit cell has eight atoms at the corners, which each contribute $\frac{1}{8}$ of an atom to the unit cell and therefore there are $8 \times \frac{1}{8}$, that is 1 atom, in each unit cell.

If there is an atom at the centre of each cube then we have a body-centred cubic (bcc) lattice. The unit cell for this structure is shown in Figure A.54.

The coordination number of each atom in a body-centred cubic lattice is eight because each atom is surrounded by eight nearest neighbours. The atom in the centre of the cube is completely within the unit cell and so contributes one atom to the unit cell. The total number of atoms in the body-centred cubic unit cell is therefore $1 + (8 \times \frac{1}{8}) = 2$.

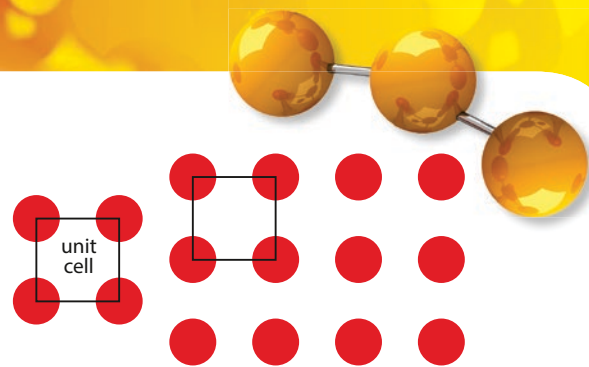


Figure A.52 A simple two-dimensional square lattice with the unit cell.

A unit cell is the simplest repeating unit from which a whole crystal can be built up.

A coordination number is the number of nearest neighbours for an atom or ion in a crystal.

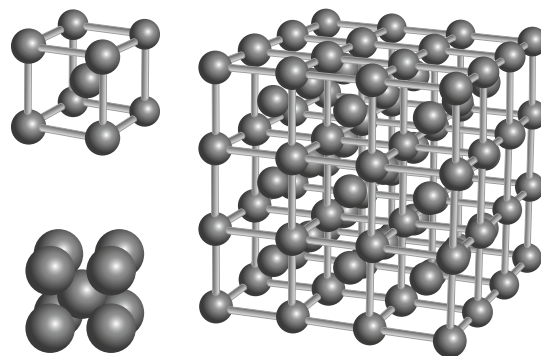


Figure A.54 A body-centred cubic structure.

fcc structures are sometimes called cubic close-packed (ccp).

If there is atom at the centre of each face then the unit cell and lattice are described as face-centred cubic (fcc).

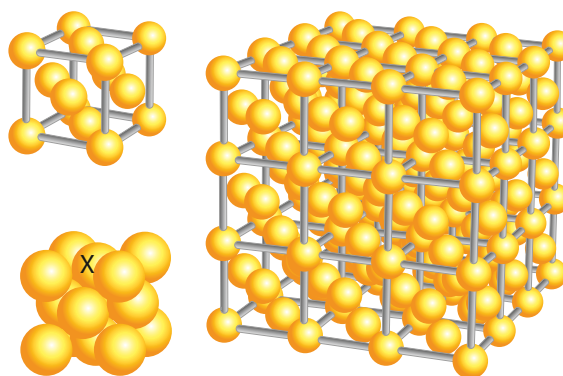


Figure A.55 A face-centred cubic structure.

Consider the atom marked 'X' in Figure A.55 – this has four nearest neighbours in the same layer, four more in the layer below and there would also be four in the layer above if another unit cell is put on top. So, the coordination number of each atom is 12.

Of the three structures discussed here, the atoms in the face-centred cubic structure have the highest coordination number and are, therefore, packed together most closely. The face-centred cubic structure is described as a **close-packed structure** in which the atoms are as close together as they can be and occupy 74.05% of the available space.

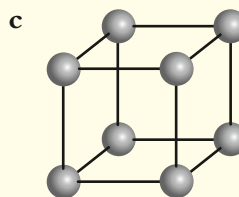
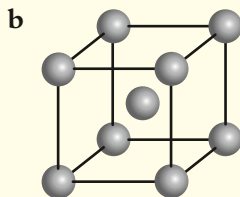
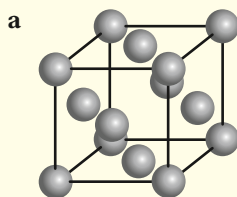
Each atom on the face of the unit cell is shared between two unit cells, so that it contributes $\frac{1}{2}$ an atom to each unit cell. The number of atoms in an fcc unit cell is therefore:

$$(8 \times \frac{1}{8}) + (6 \times \frac{1}{2}) = 4$$

There is another close-packed lattice structure adopted by metals – the hexagonal close-packed (hcp) structure.

? Test yourself

16 Identify each type of unit cell shown below and state the number of atoms in the unit cell.



X-ray diffraction

X-ray diffraction (X-ray crystallography) is a very powerful technique that can be used to analyse the structures of metallic and ionic lattices, and also determine the full structures of covalent molecules, including bond lengths and angles.

Diffraction occurs when waves spread out on passing through a gap or around a solid object. X-rays have wavelengths similar to the distances between the planes in a crystal and therefore undergo diffraction by crystals. This diffraction can be regarded as essentially the same as reflections from the crystal planes (Figure A.56a).

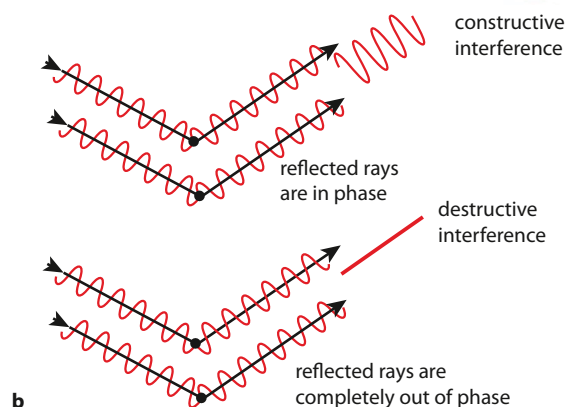
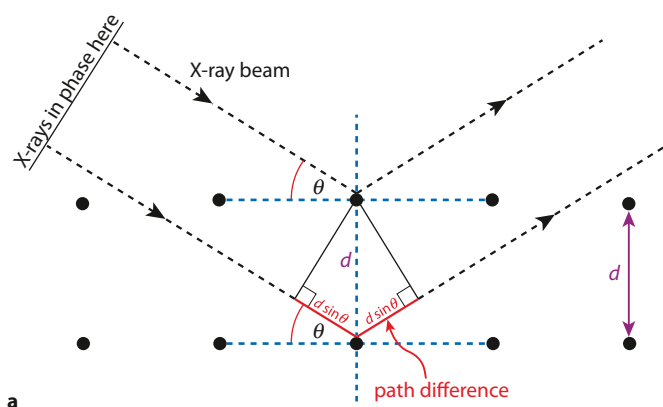


Figure A.56 **a** X-rays are reflected from planes of atoms or ions in a crystal; **b** constructive and destructive interference.

For the X-ray beams reflected from the crystal planes to give a measurable signal, constructive interference of beams from different planes must occur (Figure A.56b). The condition for this to occur is that the path difference between X-rays reflected from planes of atoms must be equal to a whole number of wavelengths so that the beams are completely in phase. This condition is summarised in the Bragg equation:

$$n\lambda = 2d \sin \theta$$

where λ is the wavelength of the X-rays, d is the distance between planes of atoms, θ is the angle at which the X-rays hit the plane of atoms and n is a whole number. When $n = 1$ we talk about a first order reflection, $n = 2$ is second order and so on.

Worked example

A.3 Polonium has a simple cubic lattice structure. When using X-rays of wavelength 1.54×10^{-10} m, the first order reflection occurs at an angle of 13.3° . Calculate: **a** the length of the unit cell; **b** the angle at which the second order reflection occurs.

- a** The lattice has a simple cubic structure, so the distance between planes of atoms, d , is equal to the length of the unit cell. The Bragg equation can be used to work out d :

$$n\lambda = 2d \sin \theta$$

Substituting the given values we get: $1 \times 1.54 \times 10^{-10} = 2 \times d \times \sin 13.3^\circ$

Rearranging and calculating, we get: $d = 3.35 \times 10^{-10}$ m, which is the length of the unit cell.

- b** Using $n = 2$ and substituting the value of d from part **a** into the Bragg equation we get:

$$2 \times 1.54 \times 10^{-10} = 2 \times 3.35 \times 10^{-10} \times \sin \theta$$

Rearranging and calculating, we get: $\sin \theta = 0.4597$

$$\text{So, } \theta = \sin^{-1} 0.4597$$

$$= 27.4^\circ$$

Determination of the density of a metal

The stages in the calculation are:

- 1 Calculate the mass of one atom of the metal by dividing the relative atomic mass by 'Avogadro's' constant.
- 2 Work out the number of atoms in the unit cell from the cell structure:
simple cubic – 1 atom per cell
bcc – 2 atoms per cell
fcc – 4 atoms per cell.
- 3 Calculate the mass of a unit cell by multiplying the mass of an atom by the number of atoms in the unit cell.
- 4 Work out the volume of the unit cell by using the cell dimensions or atomic radii.
- 5 Divide the mass of the unit cell by the unit cell volume to give the density.

Worked example

A.4 Given that the sodium has a body-centred cubic structure with unit cell dimensions of 429 pm (4.29×10^{-10} m) calculate the density of sodium metal.

The mass of a sodium atom is $\frac{22.99}{6.02 \times 10^{23}} = 3.819 \times 10^{-23}$ g

In the body-centred cubic structure there are two atoms per unit cell.

So the total mass of a unit cell is $2 \times 3.819 \times 10^{-23} = 7.638 \times 10^{-23}$ g

The volume of a unit cell is $(4.29 \times 10^{-10})^3 = 7.90 \times 10^{-29}$ m³

$$\text{Density} = \frac{\text{mass}}{\text{volume}}$$

$$\begin{aligned}\text{So the density of sodium} &= \frac{7.638 \times 10^{-23}}{7.90 \times 10^{-29}} \\ &= 9.67 \times 10^5 \text{ g m}^{-3}\end{aligned}$$

This can be converted into g cm⁻³ by dividing by 10⁶ (there are 10⁶ cm³ in 1 m³).

Therefore the density of sodium is 0.967 g cm⁻³.

To do this calculation using atomic radii instead of the unit cell dimensions, we have to do a bit of trigonometry. The atoms are packed together so that along the body diagonal of the unit cell there is the equivalent of four atomic radii (Figure A.57). The length of the body diagonal in a cube is $a\sqrt{3}$, where a is the length of a side of the cube.

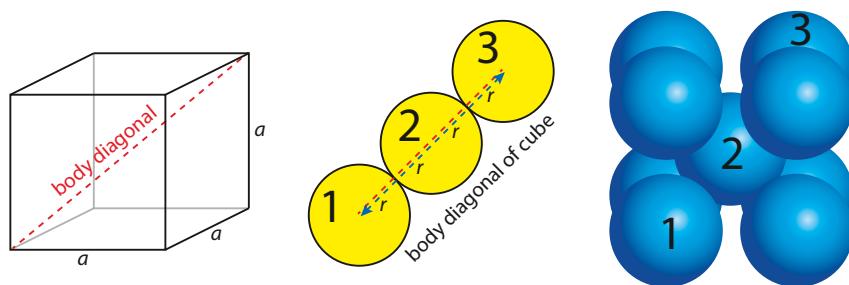


Figure A.57 Body-centred cubic unit cell.

The length of one side of the unit cell is, therefore, given by the equation $a\sqrt{3} = 4r$, where a is the length of the side and r is the atomic radius. This gives the relationship $a = \frac{4r}{\sqrt{3}}$. This can then be used in the calculation to work out the volume of the unit cell (a^3). So the volume of the unit cell is $\frac{64r^3}{3\sqrt{3}}$.

For a simple cubic unit cell (Figure A.58a), the calculation is much simpler and the length of one side of the unit cell is equal to two atomic radii. So the volume of one unit cell is $8r^3$.

For a face-centred cubic structure we have the situation shown in Figure A.58b. The length of a face diagonal of a cube of side a is $a\sqrt{2}$. This is equivalent to four atomic radii and therefore we can write $a\sqrt{2} = 4r$, which can be rearranged to give $a = 2\sqrt{2}r$. So the volume of the unit cell is $16\sqrt{2}r^3$.

Nature of science

Data are extremely important in science and they allow scientists to develop theories. X-ray crystallography is one of the most important techniques for gathering information about the structure of molecules. The double helical structure of DNA was proposed after examining X-ray data.

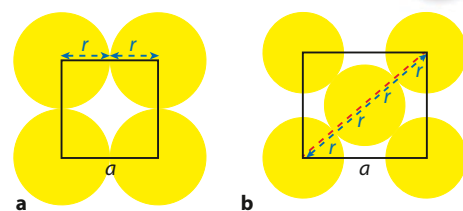


Figure A.58 a Simple cubic unit cell; b face-centred cubic unit cell.

Many high temperature superconductors have structures based on the perovskite crystal structure and can be analysed using X-ray crystallography.

? Test yourself

- 17 Calculate the angles at which constructive interference of X-rays occurs when X-rays of wavelength 1.54×10^{-10} m are incident on a crystal in which the spacing between the layers of atoms is 3.00×10^{-10} m.
- 18 Calculate the length of the unit cell for each of the following:
 - a gold: face-centred cubic lattice, atomic radius = 1.442×10^{-10} m
 - b molybdenum: body-centred cubic lattice, atomic radius = 146 pm
- 19 Calculate the mass of:
 - a a gold atom
 - b a molybdenum atom
- 20 Using data from questions 18 and 19, calculate the unit cell mass for:
 - a gold
 - b molybdenum
- 21 Using data from the questions above, calculate the density (g cm^{-3}) of:
 - a gold
 - b molybdenum
- 22 Calcium has an atomic radius of 197 pm and a face-centred cubic lattice structure. Calculate the length of one side of the unit cell, and hence the density of calcium in g cm^{-3} .

Learning objectives

- Understand what is meant by condensation polymerisation
- Describe the differences between addition polymerisation and condensation polymerisation
- Write equations for the formation of polyesters and polyamides
- Describe the structure of Kevlar[®] and explain why it is strong and also soluble in concentrated sulfuric acid

The most common molecule formed in condensation polymerisation is water. Hydrogen chloride can also be formed, depending on the starting materials – ammonia is formed very rarely.

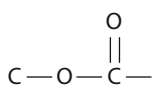


Figure A.61 The ester functional group.

A9 Condensation polymers (HL)

Condensation polymers are formed when monomers, each containing two functional groups, join together with the elimination of a small molecule such as water or hydrogen chloride.

Polyesters

We have already seen the condensation reaction of a carboxylic acid with an alcohol to form an ester in Topic 10. When the two molecules join together (Figure A.59), a water molecule is eliminated.

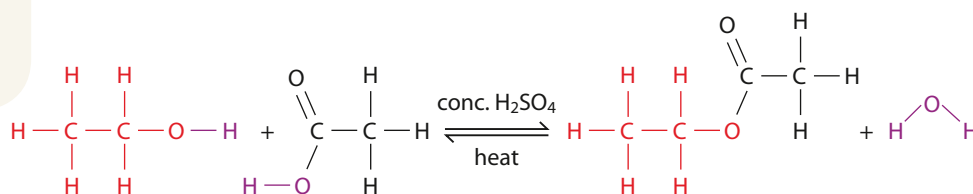


Figure A.59 A condensation reaction.

Polyesters may be formed in a **condensation** polymerisation reaction when a dicarboxylic acid reacts with a dihydric alcohol (an alcohol with two OH groups). It is the presence of two functional groups on each monomer that allows the production of a polymer chain, because an ester group is formed on both sides of both monomers.

To form a condensation polymer, two functional groups are required on each monomer.

The reaction scheme in Figure A.60 shows a representation of the reaction of two dicarboxylic acid molecules with two dihydric alcohol molecules.

The functional group joining the monomers together is the ester functional group (Figure A.61), so this is the beginning of a **polyester** chain.

The chain can continue on both sides, because the two functional groups in the original monomers means that there will either be a ‘free’ alcohol group or a ‘free’ carboxylic acid group on each end of the chain.

It can be seen from this reaction scheme that when four monomer molecules join together, three water molecules are produced.

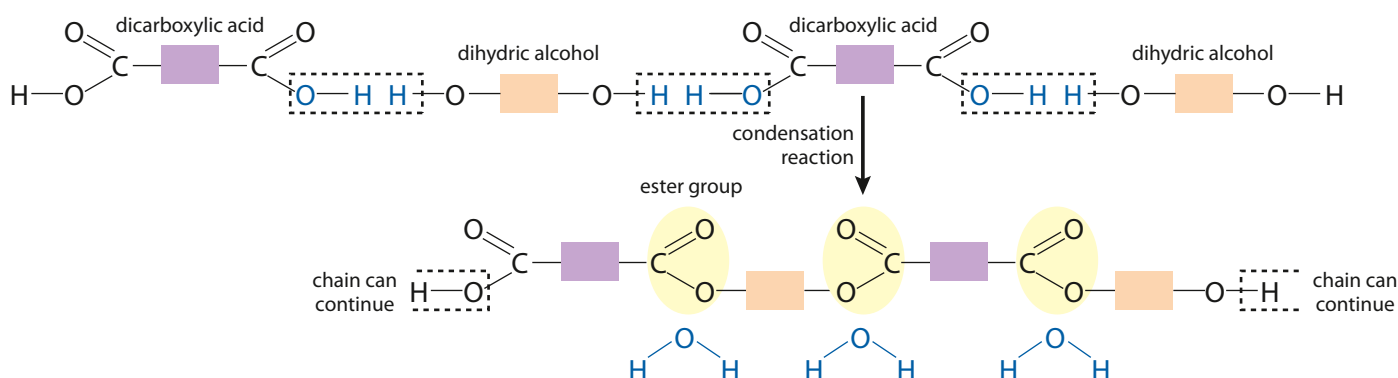


Figure A.60 Dicarboxylic acid molecules and dihydric alcohol molecules combine together and begin a polymer chain.

The total number of water molecules is always one fewer than the total number of monomer molecules that join together – as shown in Figure A.62. The polymer chain as a whole can be represented by the unit shown in brackets in the equation in Figure A.62 – this is the **repeating unit** (or repeat unit) of the polymer. The whole polymer chain could be built up by just joining these units together – a longer section of the chain is shown in Figure A.63.

In general, the repeating unit for a condensation polymer may be identified as shown in Figure A.64.

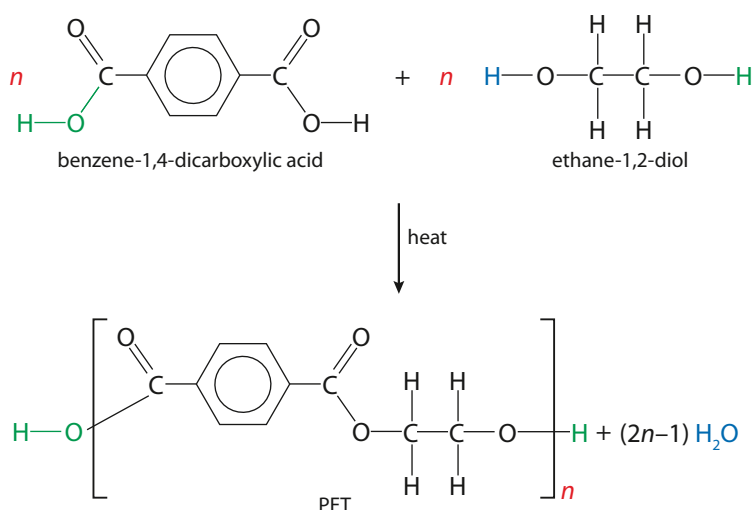


Figure A.62 Condensation polymerisation.

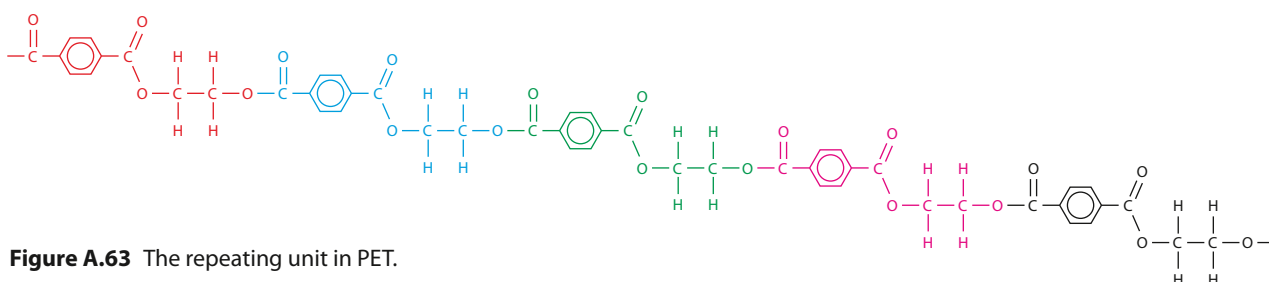


Figure A.63 The repeating unit in PET.

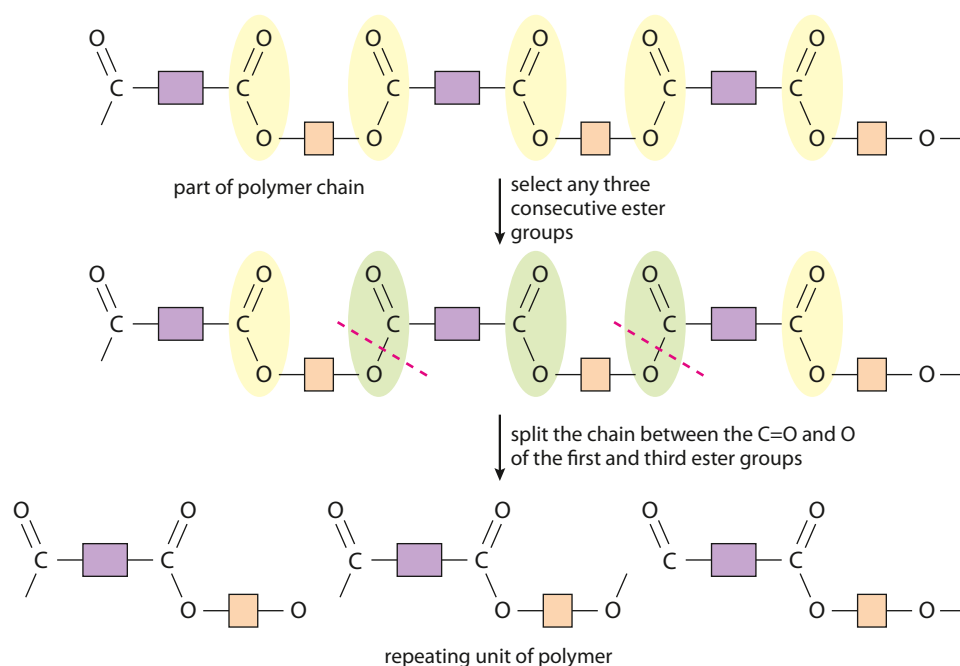


Figure A.64 Identifying the repeating unit in a condensation polymer.

The polyester formed from ethane-1,2-diol and benzene-1,4-dicarboxylic acid is commonly called polyethylene terephthalate, or PET (or PETE). PET is used in the manufacture of plastic bottles for drinks and fibres for clothing. Increasingly, PET bottles are being recycled to reduce waste.

A procedure for working out which monomers are used in making a polymer is shown in Figure A.65.

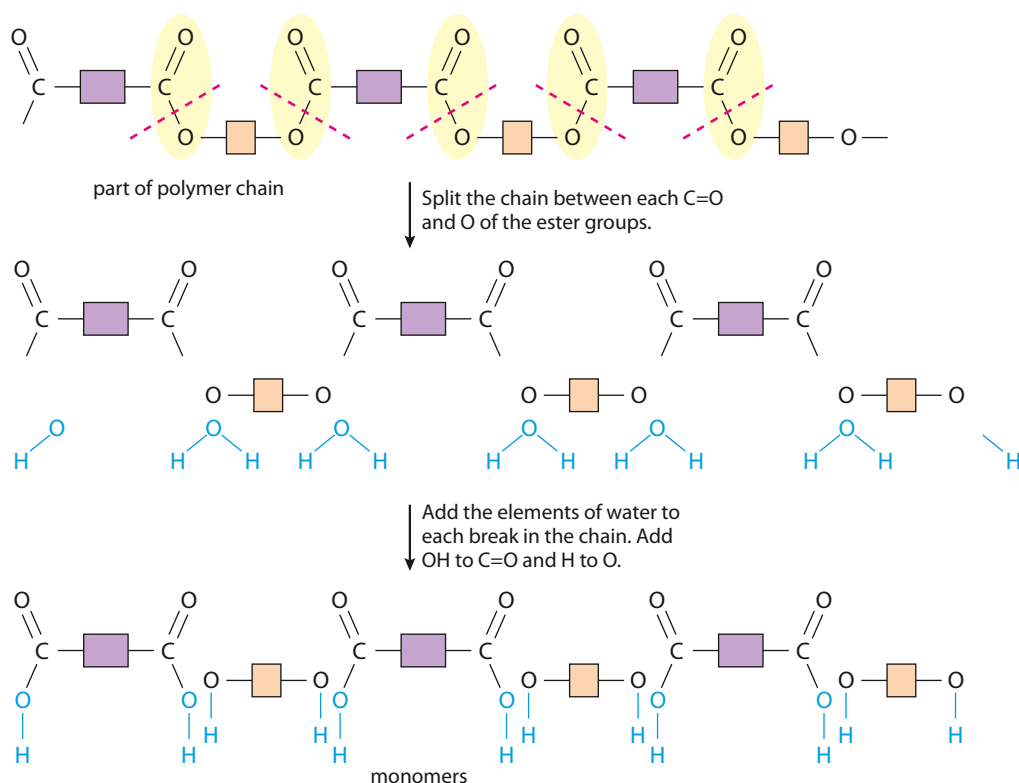


Figure A.65 Identifying the monomers used to make a condensation polymer.

Polyamides

Just as a dicarboxylic acid reacts with a dihydric alcohol to form a polyester, a dicarboxylic acid reacts with a diamine (two NH_2 groups) to form a polyamide. This is also condensation polymerisation because a water molecule is eliminated every time two monomers are joined together. A general scheme for the polymerisation reaction showing the formation of the repeating unit is shown in Figure A.66.

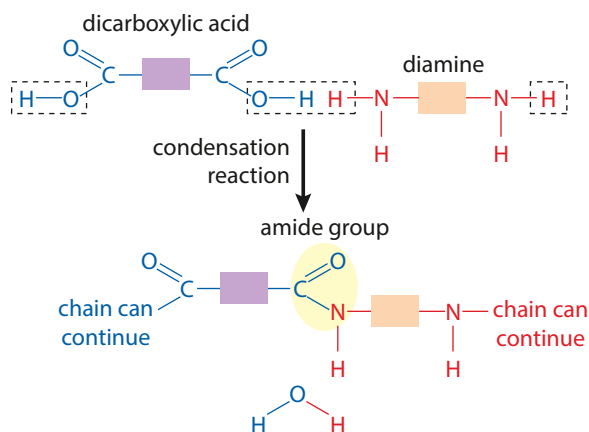


Figure A.66 Formation of a polyamide.

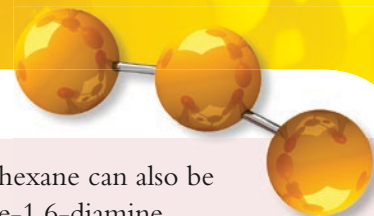


Figure A.67 shows an example of a reaction that makes a polyamide.

The polyamide formed from 1,6-diaminohexane and hexanedioic acid is commonly called 'nylon 6,6' – or 'nylon 66'. This is used in the manufacture of car parts, fibres for clothing and carpets, and some types of rope.

1,6-diaminohexane can also be called hexane-1,6-diamine.

The '6, 6' refers to the number of carbon atoms in each monomer.

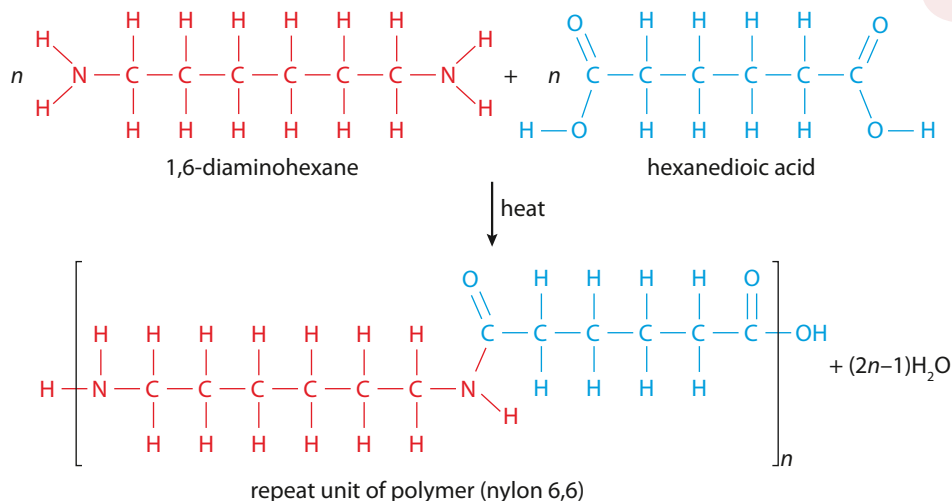
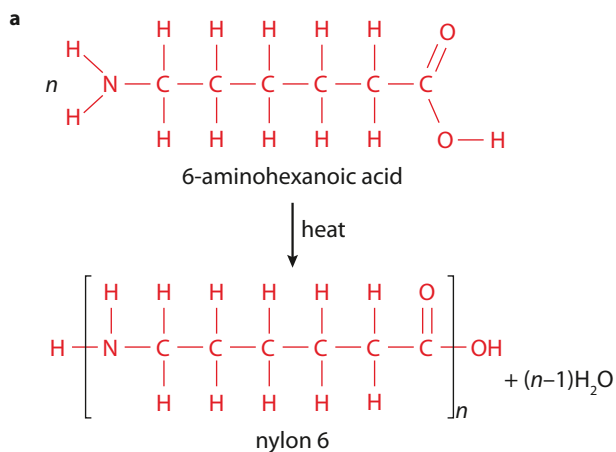


Figure A.67 Making nylon – a polyamide.

The repeating units and monomers for polyamides may be worked out in basically the same way as described in Figures A.64 and A.65.

A polyamide can also be formed by one monomer that has two different functional groups – Figure A.68a shows an example.



This is not actually how nylon 6 is produced commercially.

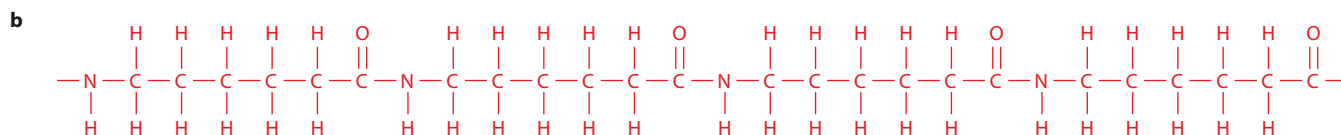


Figure A.68 **a** Formation of a polyamide; **b** part of the chain that results.

Kevlar is sometimes described as an aramid polymer – ‘aramid’ is short for ‘aromatic amide’.

Terephthaloyl chloride is an acyl chloride – it has a similar structure to a carboxylic acid, except that the OH group has been replaced with a Cl. This means that HCl, rather than water, is eliminated in the formation of Kevlar.

Kevlar®

Kevlar is a polyamide made by the reaction between 1,4-diaminobenzene and terephthaloyl chloride (benzene-1,4-dicarbonyl dichloride) (Figure A.69).

The polymer chains align themselves in an ordered way that allows for the formation of comparatively strong intermolecular hydrogen bonds between amide groups (C=O to H–N) along the whole chain (Figure A.70).

Additional intermolecular interactions arise in the form of so-called π -stacking (or π - π stacking) attractive interactions between benzene rings on adjacent chains and London forces. These hold the polymer chains together very tightly and contribute to Kevlar’s very high tensile strength.

Kevlar is a polyamide that is very strong because it has an ordered structure with relatively strong interactions between the polymer chains.

Kevlar is expensive and dangerous to manufacture because the only effective solvent for it is concentrated sulfuric acid. The formation of hydrogen bonds between sulfuric acid and the CONH groups causes the hydrogen bonds between the chains to be broken so that the Kevlar dissolves (Figure A.71).

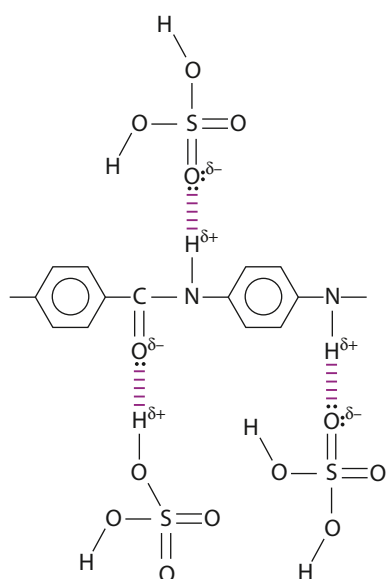


Figure A.71 Hydrogen bonding between Kevlar and sulfuric acid.

Kevlar is used in protective clothing, including body armour, synthetic ropes and sporting equipment.

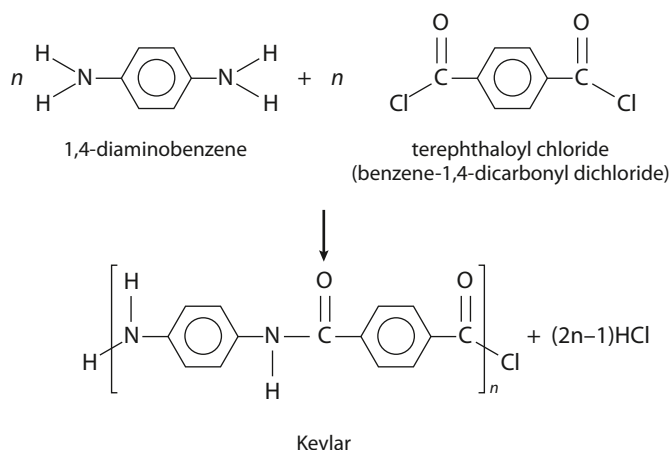


Figure A.69 Making Kevlar.

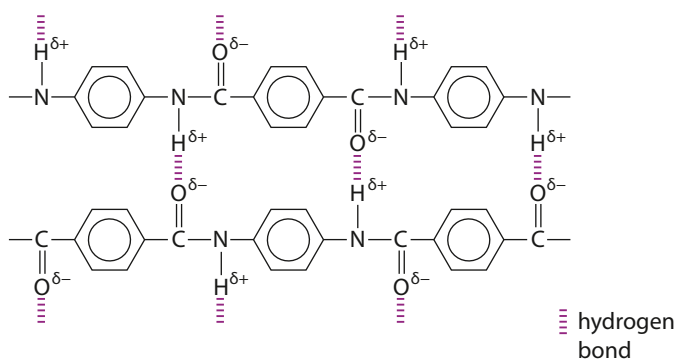


Figure A.70 Hydrogen bonding in Kevlar.

Addition	Condensation
Monomers are alkenes – the molecules used must contain C=C bonds.	Monomers are not alkenes – monomers can have different functional groups, such as COOH, OH and NH ₂ .
Monomers join together to form a long chain without the loss of anything.	Each time two monomers join together a small molecule such as water is eliminated.
Empirical formula of the polymer is the same as that of the monomer (ignoring end groups).	Empirical formula of the polymer is not the same as those of the monomers.
Typically only one monomer is used.	Typically two different monomers are used.
Monomer needs only one functional group.	Monomers must contain two functional groups.
Polymer contains mostly non-polar groups and strong bonds and is therefore chemically inert.	Polymer contains polar groups – the ester and amide groups can be hydrolysed by, for example, acids to reform the monomers.
Polymer is not biodegradable.	May be biodegradable because of the ester and amide groups between the monomers.

Table A.4 A comparison of different types of polymers.

The differences between addition and condensation polymers

Table A.4 gives a comparison of the addition polymers we discussed in Topic 10 and the condensation polymers covered in this Option.

The application of green chemistry to polymers

Green chemistry principles can be applied to the production of polymers when making decisions such as the source of the monomers used and the biodegradability of polymers.

As mentioned above, plastics derived from alkenes, such as polyethene, polypropene and polychloroethene are non-biodegradable. One solution to the problems posed by disposal of plastics is to develop biodegradable/compostable plastics.

Starch (a natural polymer of glucose molecules) has been used widely in the development of bioplastics (plastics from renewable materials) and biodegradable/compostable plastics. Examples of starch-based bioplastics include thermoplastic starch and polylactic acid (PLA).

Thermoplastic starch is obtained by mixing starch with plasticisers such as water, glycerol (propane-1,2,3-triol) and sorbitol. The plastic obtained does not have very good mechanical and physical properties and therefore it is usually blended with other polymers (either biodegradable or non-biodegradable). When blended with biodegradable polymers it can produce polymers that are fully biodegradable; when blended with non-biodegradable polymers, only the starch portion will biodegrade.

Because starch is an important energy-storage material found in plants, enzymes are present in organisms to break it down to glucose, which can be broken down further, in cellular respiration, to carbon dioxide and water. So starch is readily broken down in the environment.

Polylactic acid (PLA) is a polyester derived from lactic acid (2-hydroxypropanoic acid) (Figure A.72).

Lactic acid can be obtained from corn starch, a renewable resource, by fermentation using microorganisms. The plastic formed is biodegradable under certain conditions due to the presence of ester groups between the monomers. It has found uses in making packaging material, plastic cups etc.

The principles of green chemistry are covered in Options B and D.

‘Biodegradable’ and ‘compostable’ are not exactly the same thing – biodegradable just refers to the property that the plastic will be broken down by microorganisms such as bacteria. For a plastic to be compostable it must be broken down by microorganisms at a rate comparable with that of naturally occurring polymers such as cellulose, and not produce any toxic products.

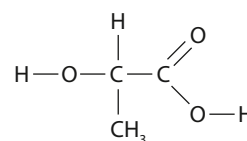


Figure A.72 Lactic acid.

Test yourself

- a**
- $$\begin{array}{ccccccc} & \text{H} & \text{H} & \text{H} & \text{H} & \text{H} & \\ & | & | & | & | & | & \\ \text{H} - & \text{O} - \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{O} - \text{H} \\ & | & | & | & | & | & \\ & \text{H} & \text{H} & \text{CH}_3 & \text{H} & \text{H} & \end{array}$$

and

$$\begin{array}{ccccccc} & \text{O} & \text{H} & \text{H} & \text{H} & \text{H} & \text{O} \\ & || & | & | & | & | & || \\ \text{HO} - & \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{OH} \\ & | & | & | & | & | & \\ & \text{H} & \text{H} & \text{H} & \text{H} & \text{H} & \end{array}$$
- b**
- $$\begin{array}{ccccccc} & \text{H} & \text{H} & \text{H} & \text{H} & & \\ & | & | & | & | & & \\ \text{H} - & \text{O} - \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{O} - \text{H} \\ & | & | & | & | & & \\ & \text{H} & \text{CH}_3 & \text{H} & \text{H} & & \end{array}$$

and

$$\begin{array}{ccccccc} & \text{O} & \text{H} & \text{CH}_3 & \text{H} & \text{H} & \text{O} \\ & || & | & | & | & | & || \\ \text{HO} - & \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{C} - & \text{OH} \\ & | & | & | & | & | & \\ & \text{H} & \text{H} & \text{H} & \text{H} & \text{CH}_3 & \end{array}$$

- a**
-
- b**
-

- $$\begin{array}{ccccccc} & \text{O} & & \text{H} & & \text{H} & & \text{O} \\ & \parallel & & | & & | & & \parallel \\ \text{H}-\text{O} & -\text{C}- & \text{C}- & \text{C}- & \text{C}- & \text{O} & -\text{H} \\ & & | & & | & & \\ & & \text{H} & & \text{H} & & \end{array}$$
 and

$$\begin{array}{ccccccc} & \text{H} & & \text{H} & & \text{CH}_3 & & \text{H} \\ & | & & | & & | & & | \\ \text{H}-\text{N} & -\text{C}- & \text{C}- & \text{C}- & \text{C}- & \text{N}- & \text{H} \\ & | & & | & & | & & | \\ & \text{H} & & \text{H} & & \text{H} & & \text{H} \end{array}$$

- O=C(Cc1ccccc1)C(=O)NCCC(C)CNCCC(C)C(=O)NCCC(C)C(=O)C2=CC=CC=C2

Nature of science

An understanding of science is essential if the public are to make informed judgments about the advantages and disadvantages of using so-called greener plastics such as PLA. It is important that scientists provide the evidence in a way that is as complete as possible but also objective so that members of the public can make their own decisions.

A10 Environmental impact – heavy metals (HL)

Heavy metals

There are many ‘heavy metals’ that are regarded as pollutants – for example lead, mercury, chromium, copper, nickel and cadmium. Rocks and minerals that contain these metals can lead to local pollution – as can mining and mineral processing. Small amounts of heavy metals may also get into the environment from various industrial sources. Table A.5 summarises some anthropogenic (from human activity) sources that may release heavy metals into the environment.

Heavy metal	Human-activity source
lead	Iron and steel production, lead water pipes. Lead was used in making paints and as a petrol additive but these are no longer permitted in most countries. There may be quite high levels of lead in soil in inner city areas where the soil has absorbed the lead emitted while leaded petrol was still in use. Some older homes may also contain lead-based paint.
chromium	industrial organic chemical industries, cement production, electroplating
mercury	waste incineration, gold mining, coal combustion, the chlor-alkali industry, inappropriate disposal of batteries, crematoria
copper	water pipes, marine paint (additives designed to control algal growth), metal-producing industries, waste incineration
cadmium	burning fossil fuels, incineration of municipal waste, smelting of zinc, lead and copper, corrosion of galvanised water pipes, electroplating, manufacture of batteries (NiCd)

Table A.5 Heavy metals and their anthropogenic sources.

Many heavy metals accumulate in the human body and can eventually lead to some serious health problems. For instance, higher levels of lead can impair the mental development of children; mercury can damage the brain, central nervous system and kidneys; cadmium can cause kidney damage, bone disease and lung and prostate cancer; chromium compounds can cause lung cancer.

Transition metals are classified as heavy metals. Certain transition metals are essential (e.g. iron in hemoglobin and cobalt in vitamin B12); however, problems arise when humans are exposed to higher levels of these metals. Transition metals can form various ions with different oxidation numbers by gaining and losing electrons, and can disturb the normal redox processes

Learning objectives

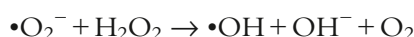
- Understand what is meant by the term ‘heavy metals’
- Understand how heavy metals can be toxic
- Compare the Haber–Weiss and Fenton reactions as possible sources of hydroxyl radicals in cells
- Explain how heavy metals can be removed from water supplies by precipitation and adsorption
- Solve problems using the solubility product constant
- Explain what is meant by polydentate ligands
- Explain the chelate effect
- Explain how chelation can be used to remove heavy metals from the human body

The term ‘heavy metal’ is very vague and imprecise and there is no clear definition of what it means. It is often used to refer to a group of metals and metalloids (such as arsenic) that have harmful environmental effects. It is also used to describe the compounds of these metals.

occurring in cells when present in higher concentrations; this is called 'oxidative stress' and it has been associated with many diseases such as cancer, Alzheimer's disease, Parkinson's disease etc.

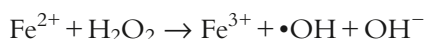
The Haber–Weiss reaction

Reactive oxygen species, such as the hydroxyl ($\text{HO}\cdot$) radical, have been identified as causes of oxidative stress in cells. The peroxide ion and hydrogen peroxide are formed in some cell processes and these could be converted into the highly reactive and potentially much more damaging hydroxyl radical by the Haber–Weiss reaction. This involves the reaction between the superoxide ion ($\cdot\text{O}_2^-$) and hydrogen peroxide to form hydroxyl free radicals:

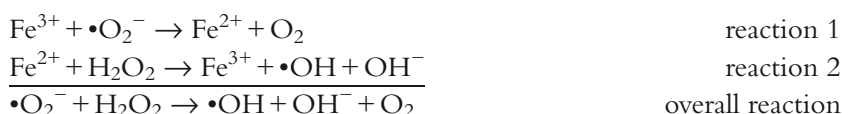


However, this reaction is extremely slow under normal conditions and its involvement in cell processes is unlikely.

It is believed that the Fenton reaction is much more likely to be involved in the production of hydroxyl radicals in cells. This reaction is between iron(II) ions and hydrogen peroxide:



Superoxide ions in a human cell can reduce Fe^{3+} ions to the +2 state (reaction 1). These iron(II) ions can then react with hydrogen peroxide in the Fenton reaction (reaction 2)



Reactions 1 and 2 taken together are equivalent to the Haber–Weiss reaction with Fe^{3+} as a catalyst – it is used up in reaction 1 and produced again in reaction 2.

The hydroxyl radical is highly reactive and interacts with many biological molecules – it damages DNA, proteins etc.

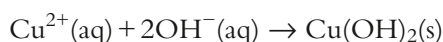
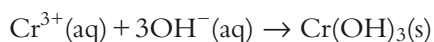
Removing heavy metals from water supplies

There are many methods that can be used to remove heavy metal ions from water supplies – we will consider precipitation and adsorption.

Precipitation

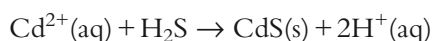
Adding various substances to water can cause heavy metal ions to form a precipitate of an insoluble compound that can be removed from the water by sedimentation and filtration.

For example, adding calcium hydroxide, $\text{Ca}(\text{OH})_2$, increases pH and removes some heavy metals as insoluble hydroxide precipitates:





Not all metals can be removed by increasing the pH. Mercury, cadmium and lead are removed by bubbling hydrogen sulfide (H_2S) gas through the water to precipitate insoluble sulfides:



The factors that influence whether a substance will precipitate or not are considered later in the section on solubility product.

Adsorption

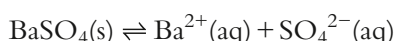
Heavy metal ions can be removed from water supplies by adsorption onto various solid surfaces. The heavy metal ions 'stick' to the surface, allowing them to be removed. Various solids can be used such as zeolites, clay minerals, metal oxides (such as TiO_2) and carbon nanotubes. For a substance to be used, it must have a high surface area on which to adsorb the heavy metal ions and be easily removed from the solution. It also helps if it is cheap and environmentally friendly.

There are several different ways in which the heavy metals can be adsorbed on to a surface. These include:

- chemisorption – formation of chemical bonds with atoms on the surface
- physisorption – formation of London forces
- ion-exchange – replacement of an ion already present in the solid with the heavy metal ion
- precipitation.

Solubility product constant

Some substances that we regard as being essentially insoluble in water are soluble to a very limited extent. In any saturated solution of a salt, an equilibrium will exist between the dissolved salt and the undissolved salt. For example, barium sulfate is commonly classed as an insoluble salt – but in reality it dissolves very slightly in water to form a dynamic equilibrium:

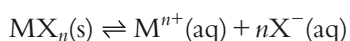


This is a heterogeneous equilibrium and an equilibrium constant, called the **solubility product constant**, can be derived:

$$K_{\text{sp}} = [\text{Ba}^{2+}(\text{aq})][\text{SO}_4^{2-}(\text{aq})]$$

The concentration of the solid does not appear in the equilibrium expression because it is essentially constant. Neither does the concentration of the water because it is in vast excess and is also effectively constant.

In general, for a salt MX_n , where X forms an X^- ion:



$$K_{\text{sp}} = [\text{M}^{n+}(\text{aq})][\text{X}^-(\text{aq})]^n$$

A solubility product constant can be worked out if you know the solubility of the substance.

A saturated solution is one in which the maximum amount of the solute is dissolved at that temperature.

The concentrations here are concentrations at equilibrium, that is, in a saturated solution.

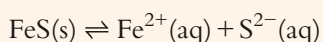
Solubility product constants are only applicable to sparingly soluble salts.

K_{sp} has no units because they are calculated in terms of activities (effective concentrations), which are relative to a standard of 1 mol dm^{-3} .

Worked example

A.5 Given that the solubility of iron(II) sulfide is $2.5 \times 10^{-9} \text{ mol dm}^{-3}$ at 298 K, calculate the solubility product constant.

The equilibrium that is established is:



The equilibrium expression is:

$$K_{\text{sp}} = [\text{Fe}^{2+}(\text{aq})][\text{S}^{2-}(\text{aq})]$$

The solubility of the iron(II) sulfide is $2.5 \times 10^{-9} \text{ mol dm}^{-3}$, so the concentration of each ion in solution will be $2.5 \times 10^{-9} \text{ mol dm}^{-3}$, and the solubility product constant can be calculated:

$$\begin{aligned} K_{\text{sp}} &= (2.5 \times 10^{-9}) \times (2.5 \times 10^{-9}) \\ &= 6.3 \times 10^{-18} \end{aligned}$$

If the product of the concentrations of the ions (using the same expression as for K_{sp}) is lower than the solubility product constant, the substance will be soluble at that temperature. If the product is greater than the solubility product constant then some solid must precipitate out of solution to bring the value back down to equal the solubility product constant. The origin of the ions does not matter.

Using FeS as an example:

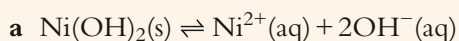
- if the product of the concentrations of Fe^{2+} and S^{2-} in a solution at 298 K is less than 6.3×10^{-18} , then all the FeS will remain in solution
- if sufficient Fe^{2+} or S^{2-} ions are added to raise the product of the concentrations above 6.3×10^{-18} , then some FeS must precipitate out of the solution.

The common ion effect

A substance AB will be less soluble in an aqueous solution containing A^+ or B^- ions than it is in water. Considering the equilibrium $\text{AB(s)} \rightleftharpoons \text{A}^+(\text{aq}) + \text{B}^-(\text{aq})$, adding $\text{A}^+(\text{aq})$ or $\text{B}^-(\text{aq})$ ions will shift the position of equilibrium to the left (Le Chatelier's principle).

Worked example

A.6 Given that the solubility product constant of Ni(OH)_2 is 6.5×10^{-18} at 298 K, calculate the solubility of nickel(II) hydroxide **a** in water; **b** in 0.10 mol dm^{-3} sodium hydroxide solution.



If the solubility of Ni(OH)_2 is represented by s , the concentration of $\text{Ni}^{2+}(\text{aq})$ in solution will be s and that of $\text{OH}^-(\text{aq})$ will be $2s$.

$$K_{\text{sp}} = [\text{Ni}^{2+}(\text{aq})][\text{OH}^-(\text{aq})]^2$$

$$6.5 \times 10^{-18} = s \times (2s)^2$$

$$6.5 \times 10^{-18} = 4s^3$$

The concentration of nickel(II) ions is the same as the solubility of nickel(II) hydroxide because for every Ni(OH)_2 unit that dissolves, one Ni^{2+} ion goes into solution.

Solving this equation:

$$s = 1.2 \times 10^{-6} \text{ mol dm}^{-3}$$

So, the solubility of nickel(II) hydroxide in water under these conditions is $1.2 \times 10^{-6} \text{ mol dm}^{-3}$

This calculation has been simplified by ignoring the hydroxide ions that come from the dissociation of water. However, the solubility is sufficiently high that this is a reasonable approximation.

- b** Sodium hydroxide is fully ionised in solution, so the concentration of OH^- ions from sodium hydroxide will be 0.10 mol dm^{-3} . The concentration of hydroxide ions is significantly higher than the solubility of $\text{Ni}(\text{OH})_2$ in water and therefore any changes in the concentration of hydroxide ions due to $\text{Ni}(\text{OH})_2$ dissolving will be negligible. We therefore assume that the overall concentration of hydroxide ions remains constant at 0.10 mol dm^{-3} . We can substitute this value into the K_{sp} expression:

$$K_{\text{sp}} = [\text{Ni}^{2+}(\text{aq})][\text{OH}^-(\text{aq})]^2$$

$$6.5 \times 10^{-18} = [\text{Ni}^{2+}(\text{aq})][0.10]^2$$

Solving this equation:

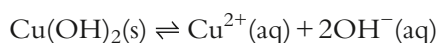
$$[\text{Ni}^{2+}(\text{aq})] = 6.5 \times 10^{-16} \text{ mol dm}^{-3}$$

The concentration of nickel(II) ions is the same as the solubility of nickel(II) hydroxide because for every $\text{Ni}(\text{OH})_2$ unit that dissolves, one Ni^{2+} ion goes into solution.

Therefore the solubility of nickel(II) hydroxide in 0.10 mol dm^{-3} sodium hydroxide is $6.5 \times 10^{-16} \text{ mol dm}^{-3}$. This is significantly lower than its solubility in pure water – the addition of a common ion (OH^-) has reduced the amount of $\text{Ni}(\text{OH})_2$ that can dissolve at 298 K.

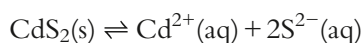
The common ion effect is used to precipitate heavy metal ions and phosphates from water that is being treated.

Because many heavy metal hydroxides have extremely small K_{sp} values, adding hydroxide ions makes even very low concentrations of these ions become insoluble, so they precipitate out of solution. Using a mixture of water and copper(II) hydroxide as an example:



If the system is at equilibrium (the solution is saturated), adding hydroxide ions to the mixture causes the position of equilibrium to shift to the left and copper(II) hydroxide precipitates out. If the solution is not saturated, hydroxide ions must be added until the product $[\text{Cu}^{2+}(\text{aq})][\text{OH}^-(\text{aq})]^2$ is bigger than K_{sp} and then copper(II) hydroxide will precipitate.

Similarly, bubbling hydrogen sulfide through the water being treated increases the concentration of sulfide ions and can cause heavy metal ions to precipitate out as sulfides, for example:



The position of equilibrium again shifts to the left as sulfide ions are added.

In the latter stages of water treatment, a coagulant is added to facilitate the formation of a sludge containing the heavy metals and other insoluble substances. The sludge settles out, is separated, dried and disposed of in landfill sites. Because it is so insoluble, the sludge does not present significant toxic issues to the environment.

Worked example

- A.7 a** A body of water has a cadmium concentration of $1.2 \times 10^{-15} \text{ mol dm}^{-3}$. Hydrogen sulfide is bubbled into the water to raise the concentration of sulfide ions to $5.6 \times 10^{-15} \text{ mol dm}^{-3}$. Given that the solubility product constant for CdS is 1.6×10^{-28} at 298 K, determine if any cadmium sulfide will precipitate out.
- b** More hydrogen sulfide is bubbled into the water until the concentration of sulfide ions is increased to $1.0 \times 10^{-9} \text{ mol dm}^{-3}$. Determine the mass of cadmium sulfide that will be precipitated from $1.0 \times 10^6 \text{ dm}^3$ of the water. (M_r of cadmium sulfide = 144.48)

a $K_{sp} = [\text{Cd}^{2+}(\text{aq})][\text{S}^{2-}(\text{aq})]$

Working out the product $[\text{Cd}^{2+}(\text{aq})][\text{S}^{2-}(\text{aq})]$:

$$(1.2 \times 10^{-15}) \times (5.6 \times 10^{-15}) = 6.7 \times 10^{-30}$$

This is lower than the solubility product constant, so all the cadmium sulfide will be soluble and none will precipitate.

- b** We can substitute values into the K_{sp} expression:

$$K_{sp} = [\text{Cd}^{2+}(\text{aq})][\text{S}^{2-}(\text{aq})]$$

$$1.6 \times 10^{-28} = [\text{Cd}^{2+}(\text{aq})] \times 1.0 \times 10^{-9}$$

Solving this equation:

$$[\text{Cd}^{2+}(\text{aq})] = 1.6 \times 10^{-19} \text{ mol dm}^{-3}$$

The concentration of cadmium ions that can be present in solution has reduced from $1.2 \times 10^{-15} \text{ mol dm}^{-3}$ to $1.6 \times 10^{-19} \text{ mol dm}^{-3}$. For this to happen, cadmium sulfide must precipitate out of the solution. The amount of cadmium sulfide that precipitates out is given by:

$$1.2 \times 10^{-15} - 1.6 \times 10^{-19} = 1.2 \times 10^{-15} \text{ mol dm}^{-3}$$

If the volume of water is $1.0 \times 10^6 \text{ dm}^3$ then the number of moles of cadmium sulfide that precipitate out is given by:

$$1.0 \times 10^6 \times 1.2 \times 10^{-15} = 1.2 \times 10^{-9} \text{ mol}$$

number of moles = volume \times concentration

The mass of CdS is given by:

$$1.2 \times 10^{-9} \times 144.48 = 1.7 \times 10^{-7} \text{ g}$$

mass = number of moles \times relative molecular mass

? Test yourself

28 Calculate K_{sp} for each of the following.

	Compound	Solubility / mol dm^{-3}
a	AgCl	1.3×10^{-5}
b	Fe(OH)_2	5.8×10^{-6}
c	Fe(OH)_3	9.3×10^{-11}

29 Given the solubility product constants in the table, calculate the solubility of each substance in water at 298 K.

	Compound	K_{sp}
a	PbSO_4	1.6×10^{-8}
b	Ag_2S	6.3×10^{-51}
c	Ag_3PO_4	1.8×10^{-18}
d	PbI_2	1.0×10^{-9}

30 Calculate the solubility of each of the following in 0.10 mol dm^{-3} sodium hydroxide solution.

	Compound	K_{sp}
a	Mn(OH)_2	2.0×10^{-13}
b	Cr(OH)_3	1.0×10^{-33}

31 K_{sp} of Co(OH)_2 is 2.5×10^{-16} at 25°C ; 10.0 dm^3 of water is known to contain Co^{2+} ions at a concentration of $1.2 \times 10^{-7} \text{ mol dm}^{-3}$. Solid sodium hydroxide is added gradually to raise the pH in stages from 7 to 8 then from 8 to 9, from 9 to 10 and from 10 to 11. Determine at which stage Co(OH)_2 will begin to precipitate out of the water.

32 The solubility product constant of aluminium phosphate is 9.8×10^{-21} at 298 K. Given that the concentration of phosphate ions in 1.0 dm^3 of water is $1.2 \times 10^{-11} \text{ mol dm}^{-3}$, calculate the mass of aluminium phosphate that precipitates when sufficient solid aluminium sulfate is added to the water to increase the concentration of aluminium ions to $1.0 \times 10^{-5} \text{ mol dm}^{-3}$.

Polydentate ligands

We discussed the idea that transition metals can form complexes with ligands such as water and ammonia in Topic 3. Ligands that bind to a transition metal ion using only one atom are called **monodentate** ligands (sometimes **unidentate** ligands). However, there are ligands that can bind using more than one atom and these are called **polydentate ligands**. The simplest of these polydentate ligands is ethane-1,2-diamine (1,2-diaminoethane or ethylenediamine). Ethane-1,2-diamine can bind to a transition metal ion using both its nitrogen atoms and is therefore a **bidentate** ligand (Figure A.73a). With three ethane-1,2-diamine ligands, an octahedral complex is formed (Figure A.73b).

Ethane-1,2-diamine is often given the symbol *en* in chemical equations.

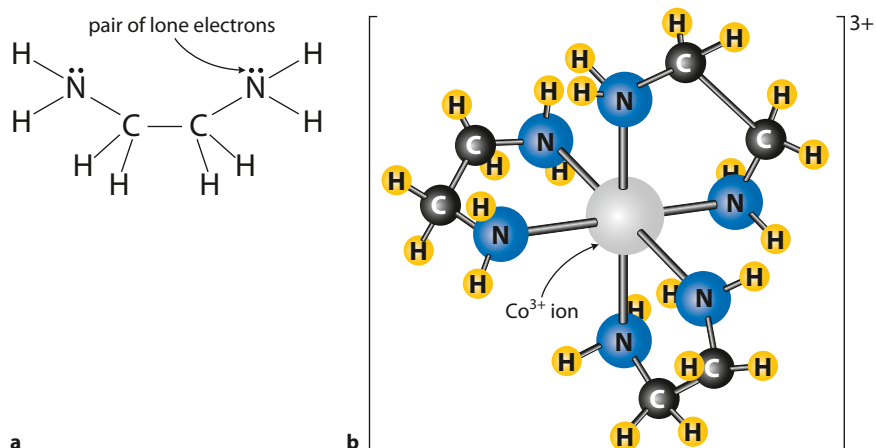
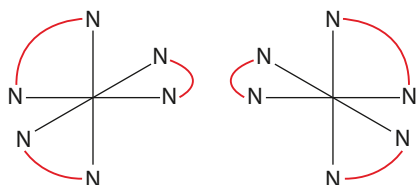


Figure A.73 Polydentate ions: **a** structure of the bidentate ligand ethane-1,2-diamine; **b** structure of the complex ion $[\text{Co}(\text{H}_2\text{NCH}_2\text{CH}_2\text{NH}_2)_3]^{3+}$.

Extension

The complex ion shown in Figure A.73b can exist as two optical isomers:



The number of atoms in a molecule that have lone pairs of electrons, and the arrangement of those atoms, will determine how a ligand binds to a transition metal ion. For instance, the ethanoate ion (Figure A.74a) can bind through both of its oxygen atoms and can act as a bidentate ligand (it can also be a monodentate ligand). But although the ethanedioate ion ($\text{C}_2\text{O}_4^{2-}$) has four oxygen atoms, the shape of the ion means that it is possible for only two of them to bond to a transition metal ion (Figure A.74b and c). (However, it can act as a tetradentate ligand when bridging between two transition metal ions.)

Figure A.75a shows the structure of ethylenediaminetetraacetic acid (this is not a systematic name but the systematic name is a real mouthful!) – commonly called EDTA. Each COOH group can lose a proton to form a $4-$ ion (Figure A.75b), which can act as a hexadentate ligand, coordinating to a transition metal ion through both nitrogen atoms and the singly bonded oxygen atoms of the carboxylate groups (Figure A.75c).

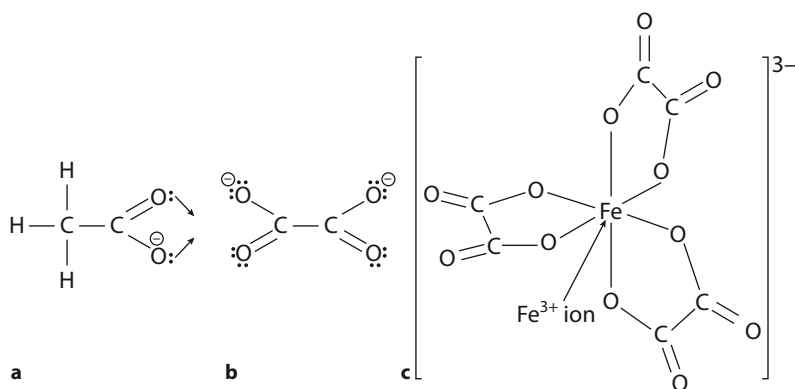


Figure A.74 **a** The ethanoate ion can be a bidentate ligand; **b** the ethanedioate ion is usually bidentate; **c** the octahedral $[\text{Fe}(\text{C}_2\text{O}_4)_3]^{3-}$ ion showing the ethanedioate ion acting as a bidentate ligand.

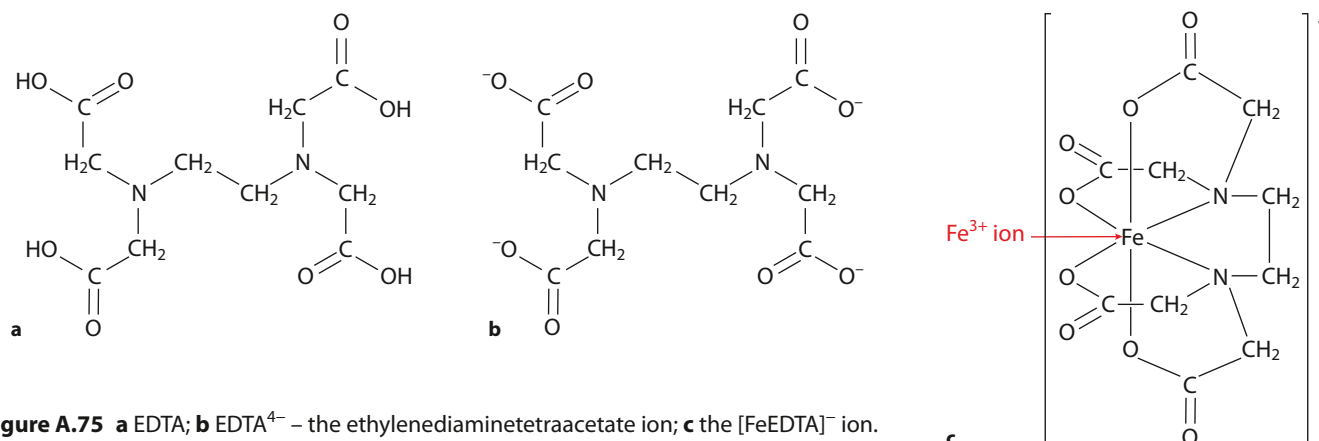
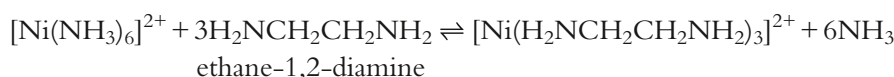


Figure A.75 **a** EDTA; **b** EDTA⁴⁻ – the ethylenediaminetetraacetate ion; **c** the [FeEDTA]⁻ ion.

The chelate effect

Polydentate ligands are also called **chelating ligands** and form **chelate complexes** (often just called **chelates**) with transition metal ions. These complexes contain a ring that includes the transition metal ion.

The ‘chelate effect’ refers to the higher stability of complexes containing polydentate ligands compared to those containing monodentate ligands with the same donor atom. Consider the following reaction:



The [Ni(H₂NCH₂CH₂NH₂)₃]²⁺ complex is more stable than the [Ni(NH₃)₆]²⁺ complex – the position of equilibrium lies more towards the right. Because the bonds formed between the Ni²⁺ ion and the nitrogen atoms are going to be very similar in strength in both complex ions, the enthalpy change for this reaction is going to be reasonably small (there will be solvent effects and so on to consider). However, the reaction involves an increase in the number of molecules from left to right and therefore an increase in entropy – it is this increase in entropy that is the driving force for the reaction. Overall, this means that formation of a chelate complex is generally a favourable process – because replacement of monodentate ligands with polydentate ones will involve an increase in entropy.

The chelate effect can be used to remove heavy metal ions from systems. Chelation may, for instance, be used to treat lead poisoning in the human body. A chelating agent (in the form of an EDTA salt) is given intravenously and this forms a chelate with the lead ions. The chelating ligand keeps the lead ions from interacting with other molecules in the body, and holds them in solution so that they can be excreted. EDTA can also bind to other heavy metal ions to form stable chelate complexes and these will also be excreted.

Nature of science

Scientific theories evolve and develop as further knowledge and evidence become available. The Fenton and Haber–Weiss reactions were discovered in the second half of the 19th century and the first half of the 20th century respectively. Interest in these was rekindled in the second half of the 20th century when the role of free radicals in oxidative stress in cells was realised. The role of the Haber–Weiss reaction in the production of hydroxyl free radicals was proposed but further evidence, in the form of kinetic studies, suggested that the reaction is far too slow to be of any significance. Further work then focused on the Fenton reaction.

Remember, $\Delta G = \Delta H - T\Delta S$

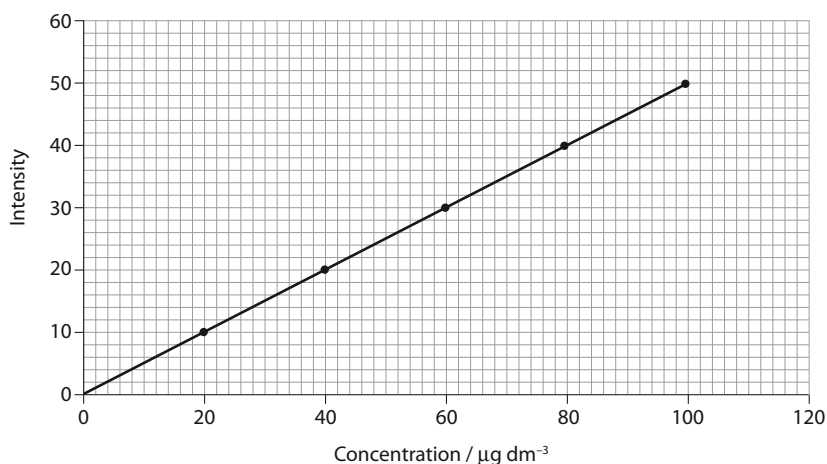
If ΔH is approximately zero, a positive value for ΔS will mean that ΔG will be negative overall and the position of equilibrium will lie more to the right.

Chelation therapy has become popular in recent years with ‘alternative’ medicine practitioners and claims have been made that it is a valid treatment for, among other things, heart disease and cancer. These claims are largely unsupported by evidence – and there are also worries that chelation therapy can be harmful.

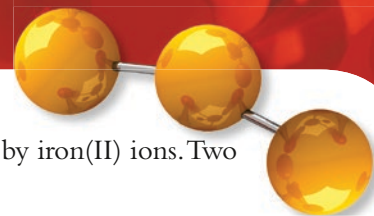
EDTA and chelating agents are used in many contexts – such as the food industry, agriculture etc. EDTA forms a stable chelate complex with metal ions and because the metal ion is surrounded by the ligand it is prevented from interacting with other species which could cause undesirable reactions.

Exam-style questions

- 1 a Give **one** example of a material in each of the following categories that is used in the construction industry:
- i metal
 - ii composite
- [2]
- b Use the bonding triangle in Figure A.4 to compare the type of bonding in MgB_2 (a superconductor) and MgO .
- [3]
- 2 Aluminium and iron are both very important construction metals.
- a Explain why iron can be extracted from iron(III) oxide by heating with carbon, but electrolysis must be used to extract aluminium.
- [2]
- b Write the half-equation for the production of aluminium at the cathode in the electrolysis of aluminium oxide dissolved in molten cryolite.
- [1]
- c A current of 1000 A is passed through a cell containing aluminium oxide dissolved in molten cryolite for 24.0 hours. Calculate the mass of aluminium produced.
- [3]
- 3 An ICP–OES experiment was carried out to determine the amount of nickel present in a sample of shellfish. 0.200 g of the shellfish flesh was taken and heated with concentrated nitric acid. The sample was made up to a total volume of 100.0 cm^3 with deionised water and analysed by ICP–OES. The intensity of the emission from the sample was 32. The calibration curve for nickel is shown below.



- a Explain how the calibration curve could be obtained.
- [2]
- b Determine the amount of nickel that would be present in 1.00 g of shellfish flesh.
- [3]

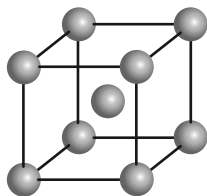


- 4 The reaction between aqueous persulfate ions ($\text{S}_2\text{O}_8^{2-}$) and aqueous iodide ions is catalysed by iron(II) ions. Two steps in the proposed reaction mechanism are shown below.
- $\text{S}_2\text{O}_8^{2-}(\text{aq}) + 2\text{Fe}^{2+}(\text{aq}) \rightarrow 2\text{SO}_4^{2-}(\text{aq}) + 2\text{Fe}^{3+}(\text{aq})$ **step 1**
- $2\text{Fe}^{3+}(\text{aq}) + 2\text{I}^{-}(\text{aq}) \rightarrow 2\text{Fe}^{2+}(\text{aq}) + \text{I}_2(\text{aq})$ **step 2**
- a Write an overall equation for this reaction [1]
- b Explain whether Fe^{2+} ions act as a homogeneous or heterogeneous catalyst in this reaction. [1]
- c Explain, using the proposed mechanism, why Fe^{2+} is described as a catalyst. [1]
- d Draw a potential energy profile for this reaction and use it to explain why a catalyst speeds up a reaction. [4]
- e Explain why zeolites can be described as selective catalysts. [3]
- f State **two** factors that must be considered when choosing a suitable catalyst for an industrial reaction. [2]
- 5 a Describe the difference between a thermotropic liquid crystal and a lyotropic liquid crystal. [2]
- b Describe, in terms of the arrangement of the molecules, the nematic liquid crystal phase. [2]
- c Describe **three** properties required by a substance to be used in a liquid crystal display. [3]
- 6 Addition polymers and plastics are highly versatile products used to make items from garden hoses to prosthetic limbs. Their versatility and wide range of uses result from different chemical structures and modification processes.
- a i What is the major structural difference between LDPE and HDPE? [1]
- ii Explain how the difference in structures affects the properties of LDPE and HDPE. [2]
- b i State **two** methods used in the modification of addition polymers. [2]
- ii For **one** of the methods in part b i, describe how it is performed and state how the properties of the product differ from the starting material. [2]
- 7 Phenol, $\text{C}_6\text{H}_5\text{OH}$ is traditionally made by the cumene process, which involves several steps but the overall reaction is:
- $\text{CH}_3\text{CH}=\text{CH}_2 + \text{C}_6\text{H}_6 + \text{O}_2 \rightarrow \text{CH}_3\text{COCH}_3 + \text{C}_6\text{H}_5\text{OH}$ **reaction 1**
- However, researchers have recently been investigating one-step syntheses, such as:
- $\text{C}_6\text{H}_6 + \text{H}_2 + \text{O}_2 \rightarrow \text{C}_6\text{H}_5\text{OH} + \text{H}_2\text{O}$ **reaction 2**
- a State the IUPAC name of all organic species in reaction 1. [4]
- b Work out the atom economy for each reaction and explain which reaction is more efficient. [4]
- 8 Carbon nanotubes have potential applications in many areas of science and technology.
- a Describe the high-pressure carbon monoxide deposition (HIPCO) method of producing carbon nanotubes. [4]
- b Write an equation for the formation of carbon nanotubes in the HIPCO process. [1]
- c i Describe the structure and bonding of single-walled carbon nanotubes. [2]
- ii Describe how the structure of a capped carbon nanotube differs from that of an open one. [1]

- 9 a Explain why plastics such as polyethene are non-biodegradable. [1]
 b Draw the repeating unit for polyethene showing all the bonds. [1]
 c One way of disposing of polymers is incineration.
 i When polyethene is burned in a limited supply of oxygen, carbon monoxide is formed. Write an equation for this combustion reaction using the repeating unit to represent the polymer. [2]
 ii Write an equation for the combustion of polychloroethene (PVC) in a good supply of oxygen. [2]
 iii Some quite complex molecules can be formed when polychloroethene is burned at lower temperatures. State the name of one class of molecules that can be formed. [1]
 iv PVC is often used for insulation around electric cables. State the name of one alternative to PVC that can be used for cabling in aeroplanes. [1]
 d An alternative to incinerating is to recycle plastics. Recycling can be a very labour-intensive process and one of the main problems is sorting the plastics into their different types. One way to distinguish between different plastics is using their infrared spectra.
 i Explain why distinguishing between HDPE and LDPE by infrared spectroscopy is difficult. [2]
 ii A particular polymer has an infrared absorbance in the range $1700\text{--}1750\text{ cm}^{-1}$. Deduce the resin identification code of this polymer. [2]

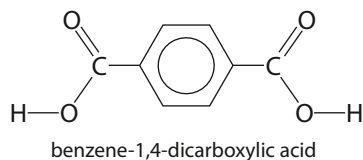
- HL 10 a State what is meant by the term 'superconductor'. [1]
 b Explain what is meant by the Meissner effect. [2]

- HL 11 a The diagram shows the unit cell of a metal crystal lattice.



- i State what is meant by the term 'unit cell'. [1]
 ii Deduce what type of lattice structure is represented in the diagram. [1]
 iii Deduce the coordination number of each atom in the lattice structure. [1]
 iv Deduce the number of atoms in the unit cell shown. [1]
 b Polonium-209 has a simple cubic lattice structure and an atomic radius of 167 pm. Calculate the density of polonium in g cm^{-3} . [6]

- HL 12 DuPont have developed a new polyester where one of the monomers (propane-1,3-diol) is made by a microorganism from corn starch rather than petroleum. The other monomer in the production of this polymer is benzene-1,4-dicarboxylic acid, the structure of which is shown below:



- a Draw the structure of propane-1,3-diol. [1]
 b Draw the repeating unit of the polymer formed from propane-1,3-diol and benzene-1,4-dicarboxylic acid. [2]
 c State **one** advantage and **one** disadvantage of making propane-1,3-diol from corn starch. [2]



- HL 13** Lead ions, Pb^{2+} , can be precipitated from polluted water by treating the water with hydrogen sulfide, H_2S . The solubility product constant (K_{sp}) of lead(II) sulfide at 298 K is 1.25×10^{-28} .
- a** Define the term 'solubility product constant' by referring to lead(II) sulfide. [1]
- b** Calculate the concentration of sulfide ions in a saturated solution of lead(II) sulfide. [2]
- HL 14 a** Explain, using an example, what is meant by the term 'bidentate ligand'. [2]
- b** With reference to the ligand that you have used in part **a**, explain what is meant by the 'chelate effect'. [3]

Option B Biochemistry

B1 Introduction to biochemistry

What is biochemistry?

Biological molecules have many functions in living organisms and these functions usually depend on the structure and shape of the molecules.

Biological molecules carry out diverse roles such as:

- the conversion of light energy to chemical energy – chlorophyll is able to do this because it has a long system of alternating single and double bonds.
- acting as catalysts in biochemical reactions – enzymes fulfil this role. They have a complex three-dimensional structure that can accommodate the bonding of other molecules both for the purpose of catalysis and for control.
- carrying genetic information – DNA molecules determine who we are. The sequence of bases in a DNA molecule constitutes the genetic code.
- providing structural support – in cellulose, for example, which is made up of long chains of sugar molecules that pack together closely to provide structure and support.
- energy production and storage – fats and sugars are involved in reactions that convert chemical energy to other forms of energy. The structure of glucose allows it to be soluble in water so that it can provide a readily available source of energy.

There are many other roles that could be listed and these and other functions of biological molecules will be discussed in subsequent sections.

Metabolism refers to the chemical reactions that go on in cells – it involves the breakdown of molecules with the release of energy and the synthesis of molecules that are required by cells.

Catabolism is the breakdown of larger molecules into smaller ones with the release of energy.

Anabolism is the process of synthesising molecules needed by cells – it requires energy.

Metabolic reactions occur in a highly controlled aqueous environment in the nucleus and cytoplasm of cells. Slight changes in pH and temperature, for example, can have large effects on the structures of biomolecules and the reactions that occur during metabolism.

An example of an anabolic process is photosynthesis. This involves using light energy from the Sun to produce energy-rich molecules (glucose) from carbon dioxide and water. The overall reaction is:



Learning objectives

- Understand that biological molecules have many functions and that these functions depend on their structure and shape
- Understand what is meant by the terms metabolism, catabolism and anabolism
- Understand that metabolic reactions take place in highly controlled aqueous environments
- Understand the difference between hydrolysis and condensation reactions
- Understand that photosynthesis involves the conversion of light energy into chemical energy
- Understand that respiration involves a complex set of reactions that provide energy for cells

An example of catabolism is the breakdown of glucose, in a series of complex steps, into carbon dioxide and water with the release of energy (in the form of ATP). This process is called respiration and the overall equation can be represented as:



This reaction provides the energy for other processes in cells.

Photosynthesis is carried out in a range of organisms from plants to bacteria and is the ultimate source of chemical energy in all the food we eat. Respiration is carried out by all organisms to produce energy from food. Photosynthesis and respiration are opposites – for every six molecules of carbon dioxide removed from the atmosphere in photosynthesis, six molecules are added by respiration. The same is true for oxygen in the opposite direction. These two processes help to maintain the balance between carbon dioxide and oxygen in the atmosphere.

Of course, there are things that disrupt this balance – for instance, burning fossil fuels, which releases carbon dioxide that was removed from the atmosphere by photosynthesis millions of years ago.

Condensation and hydrolysis reactions

These are two extremely important reactions in the build-up and breakdown of molecules in biological systems.

Condensation is the joining together of two molecules with the formation of a covalent bond and the elimination of water (Figure B.1).

Sugar molecules (to form disaccharides and polysaccharides) and amino acids (to form proteins) undergo condensation reactions.

A hydrolysis reaction is essentially the reverse of condensation – it involves breaking a covalent bond by the addition of water (Figure B.2).

Hydrolysis reactions usually occur in the presence of an acid or an alkali, or with enzymes.

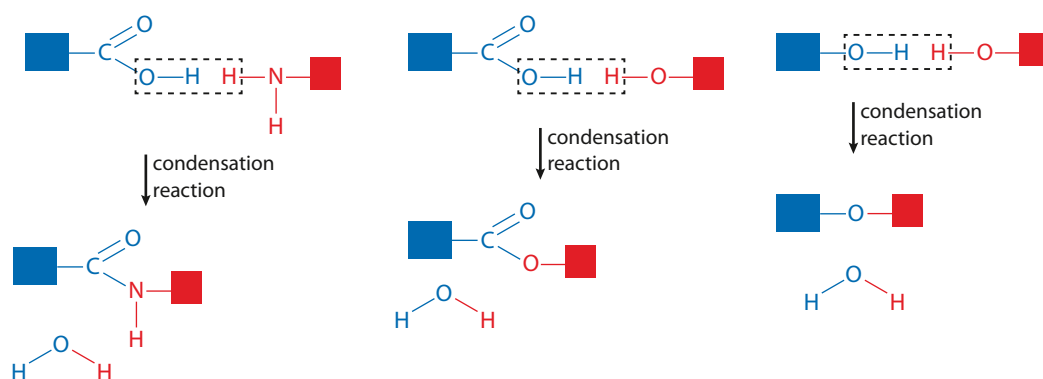


Figure B.1 Some condensation reaction types that will be met in later sections – the coloured boxes represent the rest of the molecule.

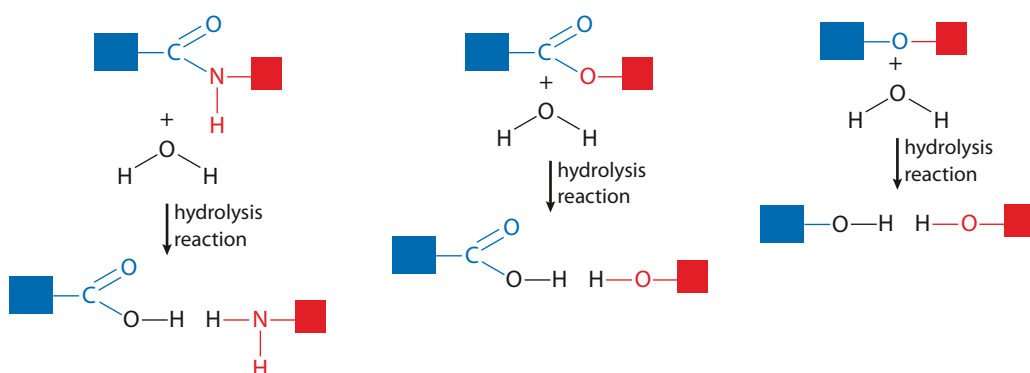


Figure B.2 Some hydrolysis reactions.

Nature of science

Scientific knowledge is ever changing and growing. The study of biochemistry is a prime example of this. It is an extremely complex field and one that is changing rapidly – a biochemistry text book written 50 years ago would be very different from one today. The double-helix structure of DNA was discovered only 60 years ago and the use of new techniques such as X-ray crystallography has allowed the elucidation of this structure as well as the structures of many other complex biological molecules.

Data is extremely important in science and the collection of large amounts of data using a variety of instruments and techniques has meant that our knowledge and understanding of the reactions that occur in metabolism has increased greatly over the last 50 years.

B2 Proteins and enzymes

B2.1 Structure of amino acids

Amino acids are the building blocks (monomers) of which proteins are made. There are 20 naturally occurring 2-amino acids that make up proteins in the body. These link together to form chains, and it is the sequence of these 2-amino acids in a chain that determines the overall structure (and therefore function) of the protein. The 2-amino acid molecules have a common structure (Figure B.3) – all have a central carbon atom to which are attached four groups:

- a carboxylic acid group – **COOH**
- an amine group – **NH₂**
- a hydrogen atom – **H**
- an **R** group – this is different in each of the 20 amino acids (see Table B.1).

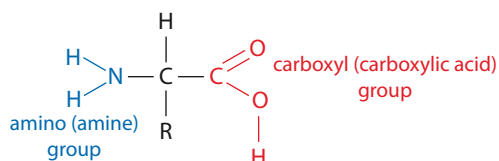


Figure B.3 General structure of a 2-amino acid.

The general chemical formula of a 2-amino acid can be written in condensed form: $\text{H}_2\text{NCH(R)COOH}$.

Learning objectives

- Understand that proteins are polymers of 2-amino acids
- Understand that amino acids are amphoteric and exist in different forms at different pH values
- Understand why amino acids have relatively high melting points and are soluble in water
- Understand the four levels of protein structure
- Understand how the function of a protein depends on its three-dimensional shape
- Understand how mixtures of amino acids and of proteins can be separated
- Understand how enzyme activity changes with substrate concentration
- Understand the effect of heavy metal ions, temperature changes and pH changes on enzyme activity

Types of amino acid

There are 20 naturally occurring 2-amino acids and, therefore, 20 different R groups. The simplest 2-amino acid is called glycine (abbreviated to Gly), where R = H. The 20 naturally occurring amino acids are shown in Table B.1.

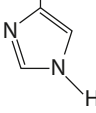
Common name	Symbol	Structural formula	pH at isoelectric point
alanine	Ala	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_3 \end{array}$	6.0
arginine	Arg	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}-\text{C}-\text{NH}_2 \\ \\ \text{NH} \end{array}$	10.8
asparagine	Asn	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{C}-\text{NH}_2 \\ \\ \text{O} \end{array}$	5.4
aspartic acid	Asp	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{COOH} \end{array}$	2.8
cysteine	Cys	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{SH} \end{array}$	4.1
glutamic acid	Glu	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{CH}_2-\text{COOH} \end{array}$	3.2
glutamine	Gln	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{CH}_2-\text{C}-\text{NH}_2 \\ \\ \text{O} \end{array}$	5.7
glycine	Gly	$\text{H}_2\text{N}-\text{CH}_2-\text{COOH}$	6.0
histidine	His	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{N} \end{array}$ 	7.6
isoleucine	Ile	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{H}_3\text{C}-\text{CH}-\text{CH}_2-\text{CH}_3 \end{array}$	6.0
leucine	Leu	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{H}_3\text{C}-\text{CH}-\text{CH}_3 \end{array}$	6.0
lysine	Lys	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{NH}_2 \end{array}$	9.7
methionine	Met	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{CH}_2-\text{S}-\text{CH}_3 \end{array}$	5.7

Table B.1 The 20 naturally occurring amino acids (continued on the next page).



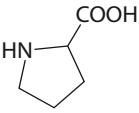
Common name	Symbol	Structural formula	pH at isoelectric point
phenylalanine	Phe	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_5 \end{array}$	5.5
proline	Pro		6.3
serine	Ser	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2-\text{OH} \end{array}$	5.7
threonine	Thr	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{H}_3\text{C}-\text{CH}-\text{OH} \end{array}$	5.6
tryptophan	Trp	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{C}_8\text{H}_6\text{N} \end{array}$	5.9
tyrosine	Tyr	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{C}_6\text{H}_4 \\ \\ \text{OH} \end{array}$	5.7
valine	Val	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{H}_3\text{C}-\text{CH}-\text{CH}_3 \end{array}$	6.0

Table B.1 (continued)

These compounds are called ‘2-amino acids’ because the amino group is on carbon 2, counting the carbon of the carboxylic acid group as carbon 1. For example, the amino acid shown in Figure **B.4a** is called 2-aminoethanoic acid; that shown in Figure **B.4b** is 2-aminopropanoic acid. Amino acids also have traditional names – 2-aminoethanoic acid and 2-aminopropanoic acid are called ‘glycine’ and ‘alanine’ respectively.

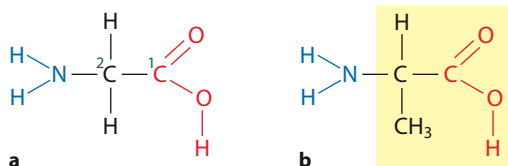


Figure B.4 **a** 2-aminoethanoic acid; **b** 2-aminopropanoic acid with the propanoic acid unit highlighted.

Extension

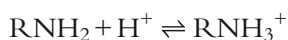
All the 2-amino acids except glycine are optically active because they have a chiral centre.

The central carbon atom is also sometimes referred to as an α -carbon, because it is next to (known as ‘alpha to’) a carboxylic acid group. This is why, in some texts, 2-amino acids are called α -amino acids.

Acid–base behaviour of 2-amino acids

Amino acids contain both **acidic** (COOH) and **basic** (NH₂) groups. So they are **amphoteric** – can act as an acid and as a base.

Amines can accept protons:



Carboxylic acids can donate protons:



The word ‘zwitterion’ comes from the German word for ‘hermaphrodite’.

The carboxylic acid group can protonate an amino group in the same molecule. When the proton is transferred from the COOH to the NH₂, a ‘neutral ion’ is formed – neutral because it has no overall charge. This form of an amino acid is known as a **zwitterion** (Figure B.5).

In aqueous solution, amino acids can exist in three different forms (Figure B.6).

Just which species dominates depends on the pH of the solution. The pH at which the concentration of a zwitterion is a maximum and the overall charge on the molecule is zero is called the **isoelectric point** and will differ depending on the R group attached to the amino acid. This difference in isoelectric points of different amino acids can be exploited analytically and it is used to separate proteins and amino acids (see electrophoresis on page 14).

If an acid is added to an amino acid at its isoelectric point, the carboxylate group (COO[−]) will be protonated – carboxylic acids are weak acids so the conjugate base is reasonably strong. The position of the equilibrium shown in Figure B.7a shifts to the right. At higher pH values than the isoelectric point a strong base can remove the extra proton from the amine group, NH₂, which is only a weak base. The position of the equilibrium shown in Figure B.7b shifts to the right.

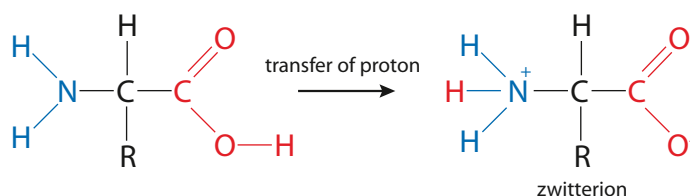


Figure B.5 Formation of a zwitterion.

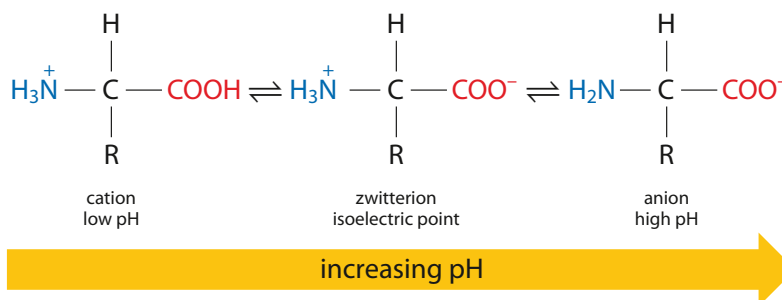


Figure B.6 Changes in charges on the amino acid as pH increases.

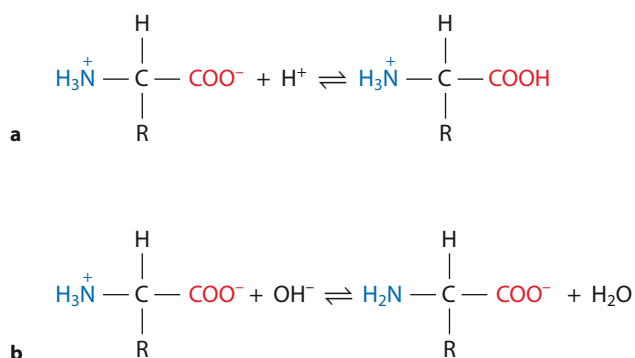


Figure B.7 **a** Protonation of a zwitterion at a pH lower than the isoelectric point; **b** deprotonation of a zwitterion at a pH higher than the isoelectric point.

Properties of amino acids

Amino acids have relatively high melting points. For example, 2-aminoethanoic acid has a melting point of about 240 °C whereas propanoic acid, which has a very similar relative molecular mass, melts at about -21 °C. These high melting points can be explained by the fact that the amino acid exists in the zwitterionic form in the solid state, and so there are stronger electrostatic attractions between oppositely charged groups on adjacent zwitterions – the COO^- group on one zwitterion attracts the NH_3^+ group on an adjacent zwitterion. This can be compared to the hydrogen bonding, dipole–dipole and London forces in propanoic acid.

Amino acids tend to be fairly soluble in water but insoluble in organic solvents – the solubility depends on the nature of the R group. The relatively high solubility in water can be explained by the fact that in aqueous solution amino acids exist as ions which can be more readily solvated (hydrated) by water molecules than neutral molecules. The solubility in water is affected by pH, with solubility generally being lowest around the isoelectric point but higher at other pH values because the amino acid then has an overall charge.

The low solubility in non-polar organic solvents can also be explained in terms of the zwitterion – the energy needed to overcome the stronger electrostatic forces between the ions cannot be paid back by the formation of relatively weak London forces between the amino acid and the organic solvent molecules.

B2.2 Structure of proteins

Proteins/polypeptides are polymers formed from amino acid monomers linked together. There are estimated to be approximately a million different proteins in the body – these differ only in the numbers and sequence of amino acids in their chains. As we shall see, the sequence of amino acids determines the overall structure (and therefore function) of a protein – it allows the protein to exist in a particular shape, this being maintained by bonds and forces between the different amino acids in the chain.

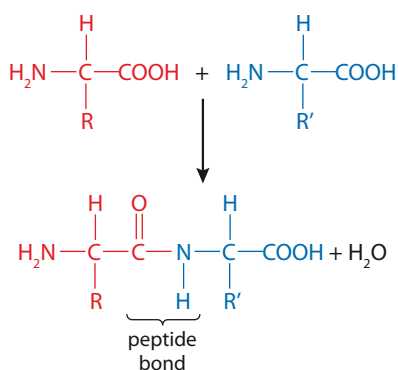
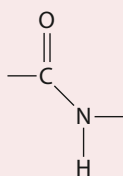


Figure B.8 Condensation reaction between two amino acids to form a dipeptide and water.

The amide functional group is:



The linear sequence of amino acids in a polypeptide chain is known as the **primary structure** of the protein, for example: Gly-Lys-Cys-Gly-Ser-Ala-Ala (glycine-lysine-cysteine-glycine-serine-alanine-alanine).

Amino acids join together to form a chain in a series of condensation reactions. The general reaction to form a dipeptide (chain of two amino acids) is shown in Figure B.8.

A covalent bond is formed when the carboxyl end of one amino acid reacts with the amino end of the other amino acid and a molecule of water is lost. The group that links the two amino acids is an amide group, and this linkage is called an **amide link**, or a **peptide bond** in proteins.

When two different 2-amino acids react together, two different dipeptides can be formed (Figure B.9).

The dipeptides are different depending on which way round the amino acids join together. The dipeptide in Figure B.9a is formed when the acid group of serine reacts with the amino group of alanine. The dipeptide in Figure B.9b is formed when the amino group of serine reacts with the acid group of alanine.

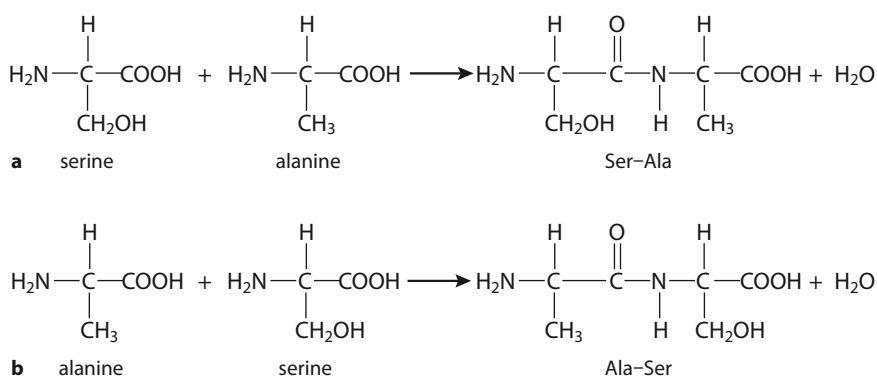


Figure B.9 Formation of different dipeptides.

The general reaction to form a chain consisting of three amino acids (a tripeptide) is shown in Figure B.10.

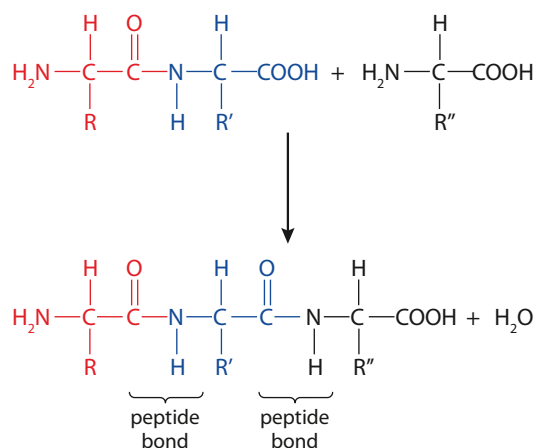


Figure B.10 Condensation reaction between a dipeptide and an amino acid to form a tripeptide and water.

When we write the sequence of amino acids in the chain, it is important to avoid confusion by everyone starting from the same end of



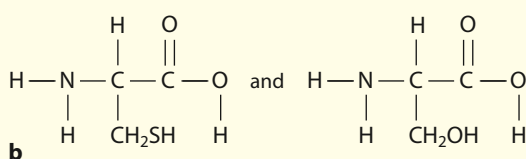
the polypeptide chain. By convention, peptide chains are always named by starting at the amino (N-terminal) end of the chain – so the sequence for the tripeptide in Figure **B.11** would be written as Gly–Cys–Ser.

Hydrolysis of proteins

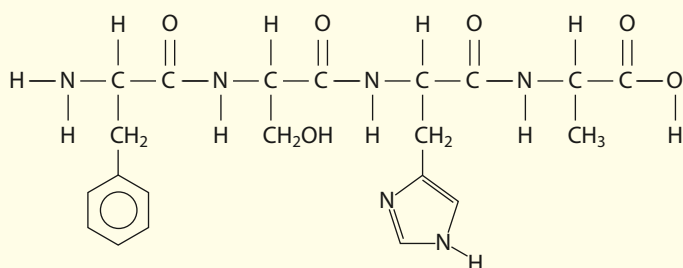
Diagram illustrating the hydrolysis of a polypeptide chain. The top structure shows a polypeptide chain with four amino acid residues (R', R'', R''', R''') and three water molecules (H₂O). The chain is broken into four fragments, each with a free carboxyl group (COOH) and a free amino group (NH₂). The bottom structure shows the resulting four amino acid fragments, each with a free carboxyl group (COOH) and a free amino group (NH₂).

Test yourself

- $$\begin{array}{c} \text{H} \quad \text{O} \\ | \quad || \\ \text{H}-\text{N}-\text{C}-\text{C}-\text{O}-\text{H} \\ | \quad | \\ \text{H} \quad \text{H} \end{array} \quad \text{and} \quad \begin{array}{c} \text{CH}_3 \quad \text{O} \\ | \quad || \\ \text{H}-\text{N}-\text{C}-\text{C}-\text{O}-\text{H} \\ | \quad | \\ \text{H} \quad \text{H} \end{array}$$



- Using Table B.1, draw the dipeptide Tyr–Trp.
- Draw the tripeptide Ala–Gly–Cys.
- Using the three-letter abbreviations for amino acids, write the sequences of all the tripeptides that can be formed from one molecule each of aspartic acid (Asp), histidine (His) and leucine (Leu).
- Use Table B.1 to name the amino acids that would be formed by the hydrolysis of the polypeptide below.



Secondary structure of proteins

Proteins do not normally exist as linear chains – they usually contain stretches in which the chain folds into regular patterns known as **α -helices** and **β -pleated sheets** (Figure B.13). This is known as the **secondary structure** of a protein.

The α -helix

The α -helix is a right-handed helix that twists in a clockwise direction from the carboxyl terminus of the polypeptide chain, with each complete turn consisting of 3.6 amino acids – it can be likened to a corkscrew. The helical structure is stabilised (held in shape) by hydrogen bonds between the C=O of one peptide bond and the NH of the peptide bond four amino acids further on. These hydrogen bonds are known as intramolecular hydrogen bonds because they exist between atoms within the same peptide chain.

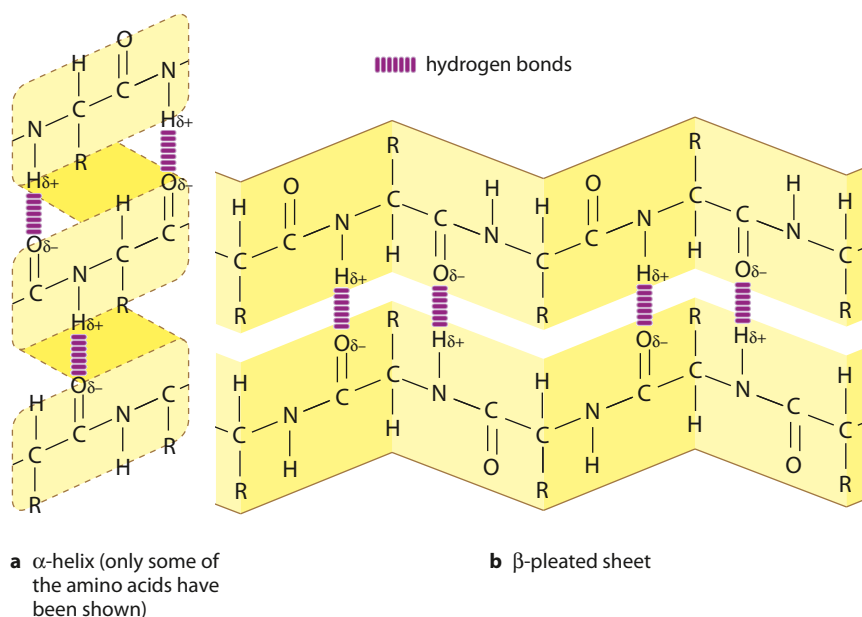


Figure B.13 a α -helix; b β -pleated sheet.



The β -pleated sheet

β -pleated sheets consist of two or more stretches of amino acids in which the polypeptide chain is almost fully extended – they take on a pleated (zig-zag) appearance, hence the name. Intramolecular hydrogen bonds form between a C=O on one strand and an NH on an adjacent strand (further on in the chain) stabilise the structure.

Tertiary structure of proteins

Every polypeptide molecule has a specific three-dimensional shape – this is known as the **tertiary structure** of the protein.

The tertiary structure exists because of a number of interactions between R groups (side chains) of amino acids in the polypeptide chain (Figure B.14) which hold the polypeptide in a particular shape. These interactions include:

- **hydrogen bonds** between amino acids bearing side chains containing, for example, OH and N
- **London forces** between amino acids bearing hydrophobic/non-polar side chains
- **electrostatic forces/ionic bonds** between, for example COO^- and NH_3^+ groups in side chains – these groups are formed when protons are transferred from an acidic group to a basic one
- **disulfide bonds** (bridges) – covalent S–S bonds formed by the oxidation of sulfhydryl (SH) groups within two cysteine residues.

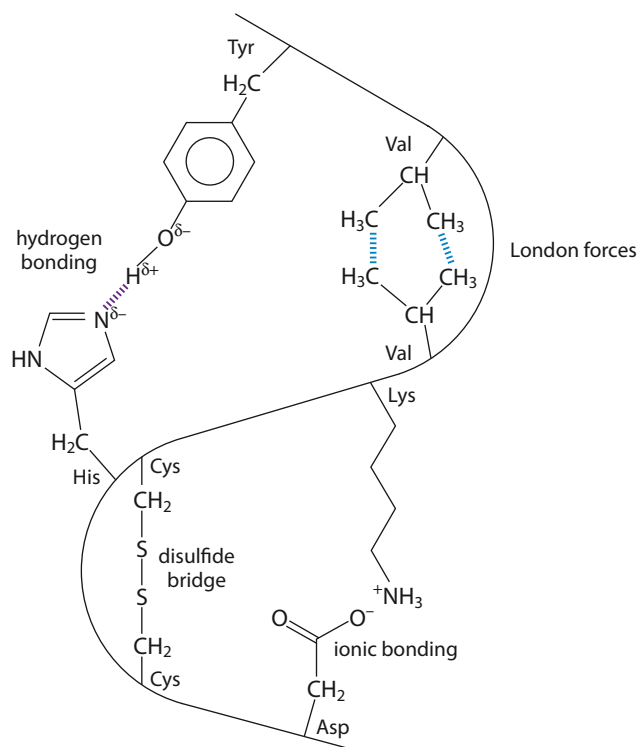


Figure B.14 The different types of interactions that can occur between some amino acids in a peptide chain.

Quaternary structure of proteins

Many proteins are made up of more than one polypeptide chain. These chains are referred to as polypeptide sub-units and associate in a specific manner – this is known as the **quaternary structure**.

The sub-units may be held together by various intermolecular interactions (interactions between groups in different protein chains), including hydrogen bonds, ionic interactions, London forces and disulfide bonds. Proteins made up of two sub-units are known as **dimers**, those with three sub-units are called **trimers** and those with four sub-units are known as **tetramers**. The sub-units may be identical to each other or may be different in structure.

Examples of tetramers containing two different types of sub-units are hemoglobin and immunoglobulin-G, an antibody with wide immunological action in the body.

Functions of proteins

Proteins serve a variety of functions in the body (see summary in Table B.2).

The function of a protein depends on its three-dimensional shape, and the shape of the protein depends on the interactions formed between the amino acids in the polypeptide chain. If these interactions and, hence, the structure are disrupted in some way – for example, changing the temperature or pH – the protein may not function properly.

Various processes operate in cells to ensure that when a protein is being made it is folded correctly – certain diseases, such as Alzheimer's disease, have been associated with protein misfolding.

Fibrous and globular proteins

Proteins can be divided into two broad classes: **globular** proteins and **fibrous** proteins.

Fibrous proteins adopt long, thin shapes and are generally insoluble in water. Fibrous proteins have a structural role and examples include keratin (in hair and nails) and collagen.

Function	Comments	Examples
structural	provide support and strength	collagen (most abundant protein in body; found in tendons, cartilage, skin, bones), keratin (found in hair and nails)
biological catalysts	enzymes catalyse biochemical reactions in the body	salivary amylase (involved in starch digestion), DNA polymerase (joins nucleotides together to form DNA) etc. (there are thousands of different enzymes in the body)
hormones	have a regulatory effect on specific cells/organs in the body	insulin (regulates blood glucose levels), growth hormone (regulates growth and cell reproduction)
immunological proteins	play a key role in the fight against infection	antibodies (recognise and bind to foreign antigens)
transport	carry materials around the body	hemoglobin (transports oxygen), serum albumin (transports many substances such as fatty acids and certain hormones)
energy source	proteins are broken down to amino acids in the body; these can enter the citric acid cycle to generate ATP (a high-energy molecule used to fuel biochemical reactions)	

Table B.2 Functions of proteins in the body.

Globular proteins are more spherical in shape, tend to be soluble in water and have roles such as catalysis (enzymes), transport etc. Examples include hemoglobin and enzymes such as pepsin.

Analysis of proteins

There are several analytical techniques that can be used to identify proteins and amino acids. Here we will focus on two: **paper chromatography** and **gel electrophoresis**.

Paper chromatography

This is a simple method for identifying the composition of amino acids in a particular protein. The protein must first be hydrolysed by heating with 6 mol dm^{-3} HCl. A small sample of the resulting mixture of amino acids is spotted onto a piece of chromatography paper and separated by suspending it in a tank containing a suitable solvent (Figure B.15).

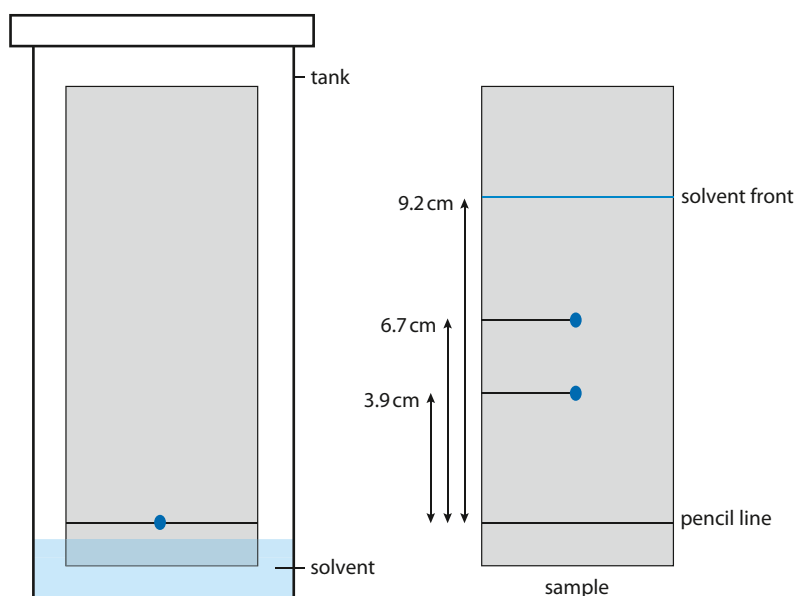


Figure B.15 Separation of amino acids using paper chromatography.

The solvent rises up the paper by capillary action and, as it does so, the amino acids also travel up the paper. The extent to which each amino acid travels up the paper depends on how it partitions between the stationary phase (the water in the chromatography paper) and the mobile phase (the solvent) – this depends on its relative solubility in each of the two phases. An amino acid that is more soluble in the stationary phase (water) than in the mobile phase (solvent) will travel less far up the paper than one that is more soluble in the mobile phase. The solubility of amino acids depends on their R groups and how they interact with solvent molecules in the stationary and mobile phases.

Once the solvent has risen to almost the top of the paper, the paper is removed from the tank. It is important to mark the distance travelled by the solvent front as soon as the paper is removed so that the **retardation factor** (R_f) can be calculated. Before the R_f values can be determined, the paper must be dried (so that the solvent evaporates) and sprayed with ninhydrin (a locating agent) – this colours the amino acids purple and allows them to be seen. Each amino acid appears as a small spot on the paper.

All chromatography techniques involve a **stationary** phase and a **mobile** phase. The components in a mixture are separated because of their differences in affinity for the stationary and mobile phases.

Exam tip

Retardation factor is sometimes called 'retention factor' in examination papers.

The next step is to measure how far each spot has travelled up the paper (the distance from the pencil line to the middle of the spot) and then divide this by the distance travelled by the solvent front (distance from original pencil line) – this calculation gives the R_f value for that particular spot:

$$R_f = \frac{\text{distance travelled by spot}}{\text{distance travelled by solvent front}}$$

Each amino acid has a characteristic R_f value when run under the same conditions and therefore can be identified by comparing the R_f value of the spot with the R_f values of known amino acids.

Exam tip

If your value for R_f is greater than 1, then you know that you have gone wrong somewhere.

R_f is less than or equal to 1 and has no units.

In Figure B.15, the distance from the pencil line to the top spot is measured and found to be 6.7 cm; the solvent front has travelled 9.2 cm from the line on which the sample was spotted. Therefore:

$$R_f = \frac{6.7}{9.2} \\ = 0.73$$

Amino acids can also be identified by spotting known amino acids on the paper alongside the unknown sample – if a particular amino acid travels the same distance up the paper as a known one, it can be identified (Figure B.16).

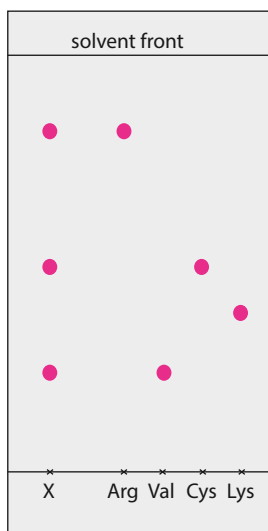


Figure B.16 The polypeptide X was hydrolysed and was found to contain arginine, valine and cysteine, but no lysine.

Gel electrophoresis

Another way of analysing amino acids and proteins is using a technique called **electrophoresis**. This technique separates charged molecules based on their ability to migrate when an electric field is applied to the system. Both proteins and amino acids can be analysed using this technique.

If the amino acid composition of a protein is to be investigated, the protein is hydrolysed using an acid (as with paper chromatography) to break the peptide bonds between the constituent amino acids. The sample mixture is then applied to a support, such as a polymer gel (poly(acrylamide) is the most common – the technique is often called poly(acrylamide) gel electrophoresis or PAGE), which is saturated with a buffer of a certain pH, used as the conducting liquid. An electric field is applied across the gel and those amino acids with negative charges at the buffer pH migrate to the positive electrode (anode), whereas those with a positive charge migrate to the negative electrode (cathode). Those amino acids with no net charge remain stationary. The distance moved by an amino acid in a particular direction depends on its size/mass and how easily it can move through the cross-linked gel. Detection of the amino acids is usually by staining, e.g. with ninhydrin.

The charge that an amino acid has at a particular pH is determined by its isoelectric point. We have already seen that the isoelectric point is the pH at which the amino acid exists in the neutral (zwitterion) form (see page 6). Different amino acids have different isoelectric points, depending on the R group attached to the central carbon. If you put an amino acid

in a solution at a pH above its isoelectric point, the amino acid will carry a net negative charge and move towards the positive electrode; if the amino acid is in a solution with a pH below its isoelectric point, it will have a net positive charge and move towards the negative electrode. Therefore, a mixture of amino acids can be separated in an electric field at a certain pH due to the differences in their isoelectric points (Figure B.17).

Mixtures of whole proteins can also be separated using gel electrophoresis. The basic set-up with a gel and a buffer is the same. The proteins move through the gel according to their charge: mass ratios – proteins with high charge or small mass move further. The charge on the protein is affected by the number of amino acids of each type (charges arise due to basic/acidic groups on side chains) and the pH of the buffer. Size and shape can also affect the migration of the protein and this will influence how easily the molecules can pass through the gel, which consists of cross-linked polymer molecules and acts like a molecular sieve. The gel must be stained to visualise the proteins – a common stain is Coomassie Brilliant Blue.

Extension

SDS–PAGE is a version of this process that separates proteins according to their relative molecular mass. A detergent – sodium dodecyl sulfate (SDS) – is added to the protein sample and this binds to most proteins in a fixed ratio. SDS partly denatures protein molecules and gives them a negative charge which is proportional to their relative molecular mass. When the treated protein mixture is run on a gel in an electrophoresis experiment, separation according to relative molecular mass will occur.

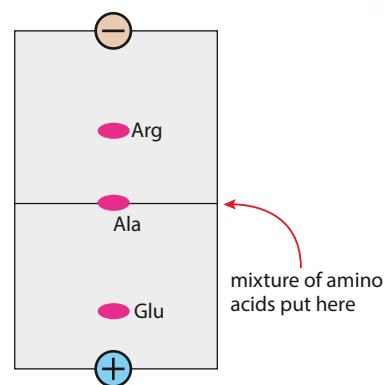
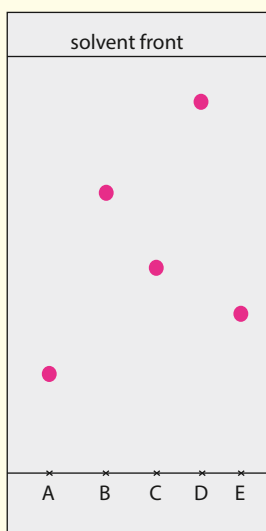


Figure B.17 In this gel electrophoresis experiment, the pH of the buffer solution was 6.0 – so the alanine did not move because its isoelectric point is 6.0. Arginine has an isoelectric point of 10.8 and so will be positively charged at pH 6.0 and move towards the negative electrode. Glutamic acid has an isoelectric point of 3.2 and so will be negatively charged at pH 6.0 and move towards the positive electrode.

? Test yourself

- 6 A paper chromatography experiment was conducted using five different amino acids, **A–E**. Calculate the R_f value of each amino acid from the chromatograph.



- 7 Use the data in Table B.1 (isoelectric points) to determine which amino acids will move towards the negative electrode in a gel electrophoresis experiment using a buffer solution of pH 8.1.

Enzymes are proteins and their activity is dependent on their shape – their **tertiary** structure.

A number of enzymes also have a **quaternary** structure, where they consist of two or more sub-units. Quaternary structure is also important for the functioning of those enzymes.

B2.3 Enzymes

Enzymes are proteins made by living organisms for the specific function of catalysing biochemical reactions. They are found throughout a cell and catalyse virtually every conversion that occurs in the cell. Some enzymes are also found outside cells – for example, in blood plasma. There are many types of enzyme and each type has a name based on the kind of reaction it catalyses and/or the kind of molecule on which it acts. For example, transferase enzymes catalyse the transfer of functional groups, oxidoreductase enzymes catalyse oxidation–reduction reactions, proteases break down proteins and lipases break down lipids. Note that the enzyme name ends in ‘-ase’ – this is true for most enzymes.

Enzymes act as biological catalysts – they speed up biochemical reactions in the body without themselves undergoing any permanent chemical change. Most reactions catalysed by enzymes would not proceed to any significant extent without the presence of enzymes – enzyme-catalysed reactions are typically 10^8 to 10^{12} times faster than the corresponding uncatalysed reactions.

They speed up reactions by lowering the activation energy of a reaction – they do not make an unfavourable reaction favourable. They do this by providing an alternative pathway for the reaction which has a lower activation energy. They also provide an area (the active site) for reactants to come together so that they are more likely to react.

The reactants involved in enzyme-catalysed reactions are called **substrates**, and enzymes are **highly specific** for the particular substrate on which they act. Some enzymes act on only one substrate, whereas others act on a group of related substrates.

An enzyme must bind temporarily to a substrate for a reaction to take place. This binding takes place in a mostly hydrophobic pocket called an **active site**. An active site is normally situated near the surface of the enzyme, towards one end of the protein. The active site is where catalysis takes place and its shape is key to the specificity of the enzyme. The types and positions of amino acids in the active site make it specific for substrates of a certain size and shape.

When a substrate (S) binds to an enzyme (E) active site, it forms a reaction intermediate known as an **enzyme–substrate complex** (ES). This is when the substrate enters the active site and forms interactions with the R groups of amino acids in the active site. This type of binding is reversible – it is relatively weak bonding such as hydrogen bonds, electrostatic interactions and London forces (not covalent bonding).

The formation of the enzyme–substrate complex can cause the substrate molecule to become strained (distorted) and this will then more readily form the product (P) along with regeneration of the enzyme. The product then diffuses away from the enzyme active site because it does not bind as effectively to the active site as the substrate does. A general equation for the reaction is:



Enzyme specificity can be compared to a key (the substrate) fitting into a lock (the active site of the enzyme). This was called the **lock-and-key hypothesis** or theory and it proposed that the active site



was a complementary shape to the substrate, like a specific key is a complementary shape to its lock (Figure B.18).

However, a more recent hypothesis proposes that when the substrate enters the active site of some enzymes it can cause a change in the shape of the active site to accommodate the substrate better – the active site changes to a shape complementary to the substrate only after the substrate has entered the active site. This is known as the **induced-fit hypothesis/theory** (Figure B.19).

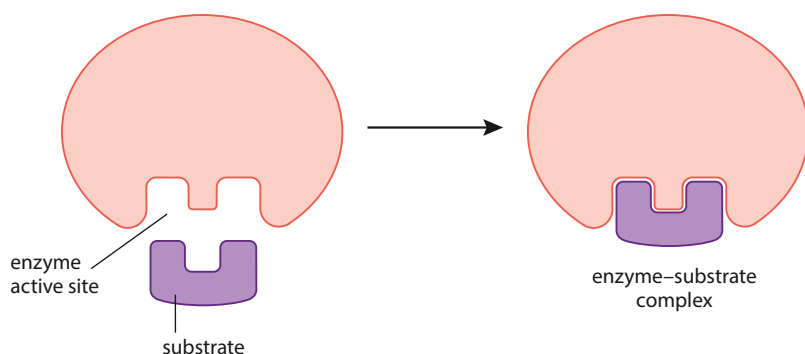


Figure B.18 The lock-and-key hypothesis.

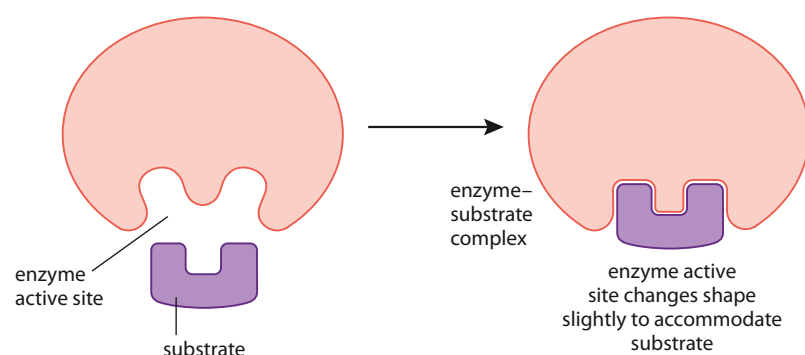


Figure B.19 The induced-fit hypothesis.

Nature of science

Science involves an ever-changing body of knowledge – the induced fit theory introduced by Koshland in the 1950s has superseded the lock and key theory.

Enzyme kinetics

At low substrate concentrations, the rate of an enzyme-catalysed reaction is proportional to the substrate concentration. However, as the substrate concentration increases, the rate of reaction increases to a lesser extent and is no longer proportional to substrate concentration. At high substrate concentrations, the rate of reaction remains constant and does not increase further with an increase in substrate concentration (Figure B.20).

When the substrate concentration is low, there is a sufficient number of enzyme active sites available to bind substrate molecules and form the enzyme-substrate complex – so, as the substrate concentration increases, more enzyme-substrate complexes form and the rate increases proportionally to the increase in substrate concentration.

As the substrate concentration increases further, however, some of the active sites are already occupied by substrate molecules and so there are

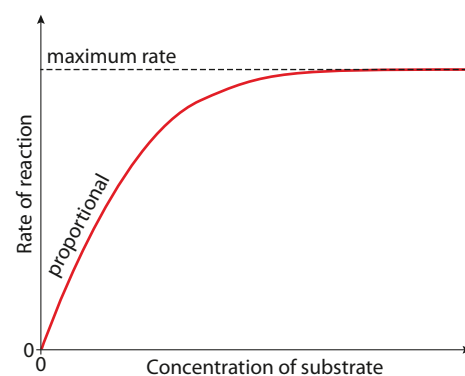


Figure B.20 How the rate of an enzyme-catalysed reaction is affected by substrate concentration. This is known as a Michaelis-Menten curve.

not enough active sites to bind all of the substrate molecules. This means that the rate of reaction increases by less than would be expected when the substrate concentration increases. This results in a slowing down in the increase of the rate of reaction and the rate is now no longer proportional to the substrate concentration.

When the substrate concentration is high, all the active sites of the enzyme molecules are occupied by substrate molecules, and so the rate of reaction remains constant (maximum rate) – any further increase in the substrate concentration results in no increase in the rate of reaction because the enzyme is saturated with substrate. The enzyme is working at maximum capacity – as soon as an active site becomes vacant, it is occupied almost immediately by another substrate molecule. The rate does not depend on the concentration of the substrate when the substrate concentration is sufficiently high.

Factors that influence enzyme activity

Enzymes have evolved to work optimally under the conditions to which they are exposed in the body. If these conditions are changed significantly, this can have a dramatic effect on the functioning of the enzyme. Three factors that can influence enzyme activity are temperature, pH and the presence of heavy metal ions.

Temperature

Temperature has an effect on enzymatic reactions (Figure B.21). As the temperature increases, the kinetic energy of the particles in the system increases and more particles have energy greater than the activation energy. There is higher chance that a substrate molecule with energy greater than the activation energy will collide with the enzyme active site, and therefore that a reaction will occur in a certain time. Also, as the temperature increases, more collisions occur in a certain time between the enzyme and substrate molecules. Therefore, increasing the temperature increases enzyme activity.

This is true only up to a certain point – if the temperature increases too much, the rate of the enzyme-catalysed reaction will decrease. This is because at higher temperatures, the vibrations of the enzyme overcome the interactions that hold the enzyme, and the active site, in its specific three-dimensional shape. As the shape of the active site is key to the activity of the enzyme, any change in shape will result in loss of function of that enzyme because it will no longer be able to bind the substrate effectively. The loss of tertiary structure is known as **denaturation**.

pH

Enzymes work within a relatively narrow pH range, depending on the pH of their environment in the body. The optimum pH varies widely from one enzyme to another – for example, digestive enzymes such as pepsin, which act in the stomach, have an optimum pH of approximately 1.5–2.5, which is the pH to which they would be exposed in the stomach. However, digestive enzymes that act in the intestines, such as pancreatic lipase, have an optimum pH of approximately 8 (Figure B.22).

If an enzyme is exposed to a pH above or below its optimum pH, activity starts to decline – this is for two reasons. Firstly, some enzymes contain ionisable amino acid side chains at the active site that participate in the catalytic action of the enzyme – changes in pH affect the ionisation

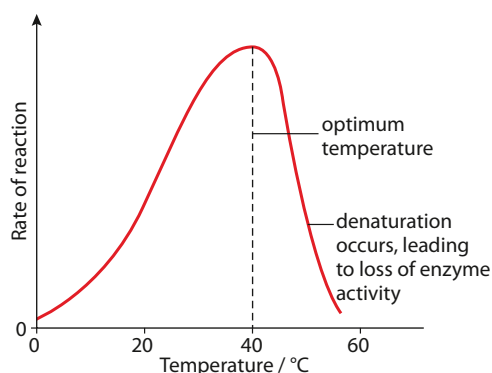


Figure B.21 The effect of temperature on enzyme activity. The optimum temperature for enzymes in the human body is around body temperature. Above that the enzyme starts to denature as bonds/forces responsible for holding the protein in its three-dimensional shape break.

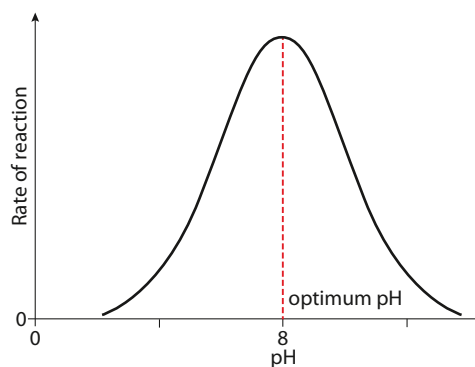
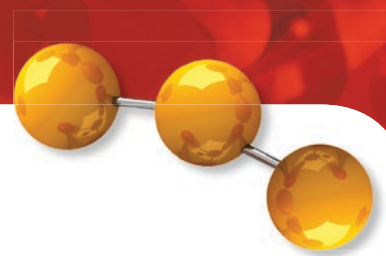


Figure B.22 Variation of enzyme activity with pH for an enzyme such as pancreatic lipase.



of those amino acids and influence their ability to participate in the reaction. For instance NH_2 groups may be protonated to NH_3^+ at lower pH values.

A second reason for the decline in enzyme activity is denaturation. Significant changes in pH alter the charges on the R groups of the amino acids that form the intramolecular interactions which dictate the shape of an active site. The resultant change in shape causes a loss of function of that enzyme, as described above for temperature.

Heavy metal ions

Heavy metals ions such as silver (Ag^+), mercury (Hg^{2+}) and lead (Pb^{2+}) have a strong affinity for sulfhydryl (SH) groups – they react with the sulfur atoms of the sulfhydryl groups found in cysteine residues in the enzyme, replacing the hydrogen atom. If these cysteine residues are involved in forming interactions that contribute to the tertiary structure of the enzyme, the shape of the enzyme is altered and it can no longer function correctly. This is the mechanism by which heavy metals are poisonous.

Nature of science

The advancement of science relies on sharing knowledge. There are now many ways that scientists can share their knowledge but the most common is through peer-reviewed journals. Scientists submit papers to a journal and these are sent to other scientists who review the work and submit recommendations as to whether it should be published or not. They consider factors such as correctness, originality and relevance. These journals are commercial enterprises, sometimes controlled by publishing houses. Over recent years there have sprung up an enormous number of journals and there have been worries from scientists about the cost of access to papers. Open-access journals (access is free) which scientists can pay to publish their papers in have also begun to appear – but there are worries in the scientific community that papers are not necessarily reviewed as thoroughly.

In the first half of the twentieth century scientists were still unsure about which substance in cells carried the genetic material – many scientists believed that it was proteins. Over a period of about 25 years from the late 1920s a series of experiments were carried out in different laboratories around the world that eventually led to the acceptance of DNA as the genetic material. This great leap in our knowledge and understanding was facilitated by the fact that scientists share their findings with each other in journals and at conferences, so that their knowledge is available for other scientists to build on.



Test yourself

- 8 Which of the following statements are true?
- a The rate of an enzyme-catalysed reaction increases with increasing temperature.
 - b The optimum pH for all enzymes is 7.
 - c Na^+ and K^+ are heavy metal ions that inhibit enzymes.
 - d The rate of an enzyme-catalysed reaction is proportional to the substrate concentration for all substrate concentrations.

Learning objectives

- Describe the uses of lipids
- Understand why fats yield more energy than carbohydrates when oxidised
- Compare lipids and carbohydrates as energy-storage molecules
- Understand that triglycerides are esters that can be formed when glycerol undergoes a condensation reaction with fatty acids
- Classify fats as saturated, monounsaturated and polyunsaturated
- Predict the melting points of fats and fatty acids based on structure
- Work out the iodine number of fats and fatty acids
- Understand the structure of phospholipids
- Work out the products of hydrolysis of triglycerides and phospholipids
- Understand hydrolytic rancidity and oxidative rancidity and the conditions that favour them
- Understand the structures of steroids
- Discuss the use and abuse of steroids
- Discuss the health effects of lipids

The term 'energy density' is used in the syllabus which is sometimes used to represent the amount of energy released per gram of fuel, but can also be used to describe the energy released per unit volume of fuel consumed.

B3 Lipids

There are various types of **lipid** in the human body – steroids (e.g. cholesterol, a steroid that is abundant in cell membranes and also is a precursor for many steroid hormones), triglycerides (fats acting as an energy store) and phospholipids (in cell membranes). Lipids, therefore, make up a broad class of molecules but they are linked by the fact that they are mostly non-polar and insoluble in water.

The uses of lipids

Some of the main uses of lipids in the human body are:

- as a structural component of cell membranes
- for energy storage
- for thermal insulation – insulates from heat loss through the skin
- for electrical insulation – myelin, which is made mostly from lipids, forms an electrically insulating layer around neurons (nerve cells)
- to provide a medium in which lipid/fat-soluble vitamins can be transported and stored
- as hormones.

Lipids and energy storage

Fat (lipid) is an important energy store in the body because it has a much higher specific energy than carbohydrate or protein. More than twice as much energy is released per gram of fat than per gram of carbohydrate or protein.

$$\text{Specific energy} = \frac{\text{energy released}}{\text{mass of substance consumed}}$$

Fats yield approximately 37 kJ of energy per gram, whereas carbohydrates and proteins each yield approximately 16 kJ per gram.

Energy is obtained from food in a process called **cellular respiration**. Food molecules, such as fats (as fatty acids) and carbohydrates (as glucose) are oxidised in a series of enzyme-catalysed reactions to ultimately produce carbon dioxide, water and energy. The carbon atoms in fatty acids are less oxidised/more reduced than in carbohydrates or proteins, and so they are able to undergo more oxidation, resulting in the release of more energy.

The average oxidation number of carbon in a fat such as $\text{C}_{57}\text{H}_{100}\text{O}_6$ is -1.54 , whereas the average oxidation number of carbon in glucose, $\text{C}_6\text{H}_{12}\text{O}_6$ is 0 . The oxidation number of carbon in CO_2 is $+4$ and so the carbon atoms in fats can be oxidised more, releasing more energy.

These values were worked out by assuming that the oxidation number of hydrogen is $+1$ and that of oxygen is -2 .

However, carbohydrates, such as glucose, have an advantage over lipids in that they are water-soluble and can therefore be transported around the body more easily. Also, the energy can be released more quickly from carbohydrates than from fats.

Lipids are insoluble in water so they are difficult to transport around the body. However, their insolubility has an advantage in that it means that they can be stored in cells without affecting the movement of water in and out of the cell by osmosis.

If glucose were present in a cell at high concentration it would cause a rapid influx of water into the cell, which could cause the cell to burst. However, glucose is not stored in cells in large quantities, rather it is converted to glycogen (a polymer of glucose) which is stored in solid form.

Triglycerides

Triglycerides are the most abundant class of lipids and make up the majority of lipids in the diet – fats and oils consist of triglycerides. Fat (as triglycerides) is found in cells known as adipocytes, which make up adipose (fatty) tissue. Adipose tissue has a number of roles in the body – it acts as a major energy reserve; it provides insulation from heat loss through the skin; and it insulates, protects and supports organs such as the heart and kidneys.

Triglycerides are non-polar, hydrophobic molecules. They are formed by condensation of three fatty acids with the three alcohol groups of glycerol (propane-1,2,3-triol) (Figure B.23). The functional group formed is an ester group and so this reaction can also be called **esterification**.

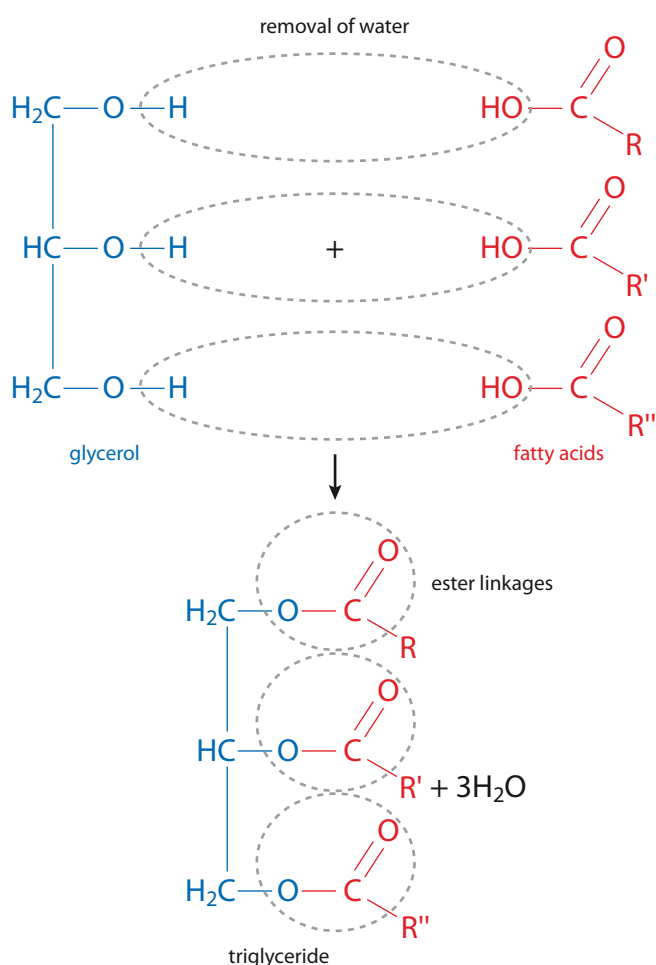


Figure B.23 Condensation of glycerol with three fatty acids to form a triglyceride, where R, R' and R'' are long-chain hydrocarbons; each individual reaction is a condensation reaction and results in the elimination of water.

Osmosis is the movement of water (or other solvents) across a semi-permeable membrane from a less concentrated solution to a more concentrated one.

Triglycerides are not usually made up of the same fatty acid – they usually contain different fatty acids.

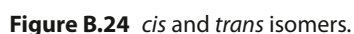
Fatty acids are molecules containing a long hydrocarbon chain with a carboxylic acid (COOH) group at one end. They have the general formula $\text{CH}_3(\text{CH}_2)_n\text{COOH}$.

There are two major types of fatty acid:

- those that have only carbon–carbon single bonds (**saturated** fatty acids)
- those that contain one or more carbon–carbon double bonds (**unsaturated** fatty acids).

Unsaturated fatty acids can exist in two forms – *cis* and *trans* (Figure **B.24**). These represent two different arrangements of the hydrogen atoms attached to the carbon atoms in the double bond. In the *cis* form the hydrogens are both on the same side of the double **bonds**, whereas in the *trans* form they are positioned on opposite sides.

The C=C bond in fatty acids is almost always *cis*, and this results in a bend (or kink) in the hydrocarbon chain (Figure **B.25**).



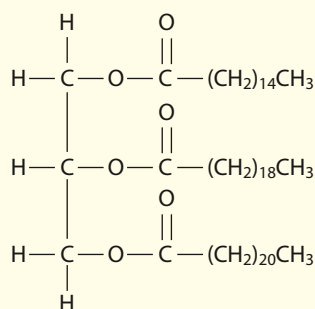
cis and *trans* isomerism is not covered in the Standard Level core syllabus. See the Higher Level section of Topic **10** on page **489** of the Coursebook for more detail on this.





? Test yourself

- 9 Draw the structural formula of the triglyceride formed when three molecules of stearic acid (Figure B.25) react with glycerol.
- 10 The formula of a triglyceride is shown below. Deduce the structural formulas of the molecules from which it was made.



Melting points of fats and oils

There are two main factors that affect the melting points of fats and oils – relative molecular mass and the degree of unsaturation.

Relative molecular mass

All other things being equal, increasing the relative molecular mass of a fatty acid will increase its melting point. Longer chain fatty acids tend to be solid at room temperature – stearic acid ($\text{C}_{18}\text{H}_{36}\text{O}_2$, M_r 284.54) melts at 69.6°C whereas lauric acid ($\text{C}_{12}\text{H}_{24}\text{O}_2$, M_r 200.36) melts at 44.2°C . The rise in melting point as the chain gets longer is due to an increase in the number and combined strength of the intermolecular London forces along the length of the chains.

Unsaturation

The level of unsaturation in the fatty acid chains influences the melting point of the fat or oil. Triglycerides with a high proportion of saturated fatty acids have higher melting points and are solids at room temperature – animal fats such as lard and butter, for example. This is because of the long saturated hydrocarbon chains being able to pack closely together, forming stronger intermolecular interactions (London forces) between the triglyceride molecules.

On the other hand, triglycerides rich in monounsaturated and polyunsaturated fatty acids have lower melting points and are oils at room temperature – like vegetable and fish oils. As we have already seen, a $\text{C}=\text{C}$ bond results in the formation of a kink in the hydrocarbon chain. As a result, the hydrocarbon chains on adjacent molecules are not able to approach each other as closely and cannot form as strong intermolecular interactions, hence the lower melting point (Figure B.26). As the number of $\text{C}=\text{C}$ bonds increases in the triglycerides they become more and more fluid so triglycerides rich in polyunsaturated fatty acids have lower melting points than those rich in monounsaturated fatty acids, although both types are liquids at room temperature. It is customary to call solid triglycerides ‘fats’ and liquid triglycerides ‘oils’.

Stearic acid and lauric acid are both saturated fatty acids.

Mono- and polyunsaturated fatty acids generally have lower melting points than saturated fatty acids of similar relative molecular masses.

The bigger the number of double bonds (the more unsaturated the triglyceride), the lower the melting point.

Most fats contain a mixture of saturated, monounsaturated and polyunsaturated fatty acids. Those fats predominantly made up of saturated fatty acids are known as saturated fats; those with predominantly monounsaturated fatty acids are called monounsaturated fats (oils); those with predominantly polyunsaturated fatty acids are polyunsaturated fats (oils).

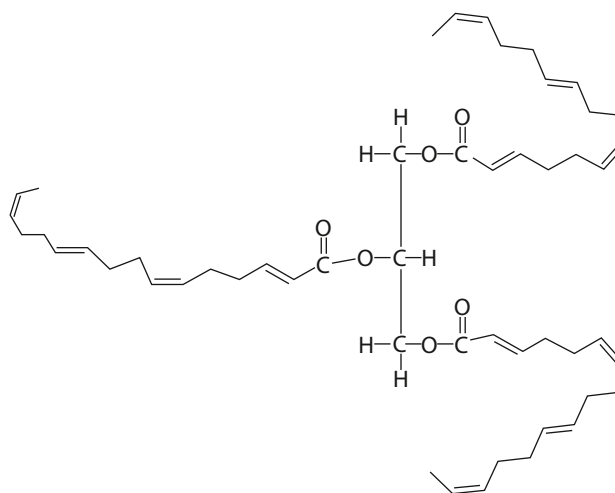


Figure B.26 The *cis* double bonds cause the chains to be kinked and molecules cannot pack together as well.

Iodine number

Using iodine it is possible to determine the **degree of unsaturation** (the number of double bonds) of a fat or an oil. Iodine can add across a double bond (Figure B.27). One molecule of iodine reacts with one C=C bond, so the number of moles of iodine used up in the reaction with a fat or oil can be equated to the number of double bonds in the fat or oil molecule. For example, if two moles of iodine reacts with one mole of fat or oil then that would indicate that there were two C=C bonds in the fat or oil molecule.

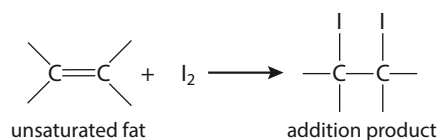


Figure B.27 The addition reaction between iodine and the double bond in an unsaturated fatty acid.

Iodine number is a measure of the degree of unsaturation in a fat or oil. It is the number of grams of iodine that reacts with 100 g of fat or oil.

A saturated fat, with wholly saturated fatty acid molecules, will have an iodine number of 0, whereas an oil consisting of mostly polyunsaturated fatty acid molecules will have a high iodine number. For example, if 65 g of iodine reacted with 50 g of a fat then the fat's iodine number would be:

$$\left(\frac{65}{50}\right) \times 100 = 130$$

An iodine number of 130 suggests that the oil contains a high degree of polyunsaturated fatty acids such as linolenic acid (which has three C=C bonds). An animal fat such as butter, with a high degree of saturated fatty acids, has an iodine number of between 25 and 45.



Worked examples

B.1 Work out the iodine number of linoleic acid.

Its structural formula is $\text{CH}_3(\text{CH}_2)_4(\text{CH}=\text{CHCH}_2)_2(\text{CH}_2)_6\text{COOH}$ and its M_r is 280.50 (M_r for $\text{I}_2 = 253.80$).

To work out the iodine number of a known fatty acid, we must count the number of $\text{C}=\text{C}$ double bonds in its molecule.

Linoleic acid has two double bonds and so 2 mol of iodine (I_2) will react with 1 mol linoleic acid. This means that 280.50 g of linoleic acid would react with 253.80×2 , or 507.60 g of iodine.

So the mass of iodine that would react with 100 g of linoleic acid is:

$$\left(\frac{507.60}{280.50}\right) \times 100 = 180.96 \text{ g}$$

The iodine number for linoleic acid is 181.

B.2 Work out the number of $\text{C}=\text{C}$ bonds in the fatty acid $\text{C}_{21}\text{H}_{37}\text{COOH}$ and calculate its iodine number. (M_r for $\text{I}_2 = 253.80$)

We saw in Topic 10 that for every two hydrogen atoms fewer than in the alkane with the same number of carbon atoms, there is one double bond or ring present. To find the number of $\text{C}=\text{C}$ bonds we need to look at the alkyl part of the molecule – in this case, $\text{C}_{21}\text{H}_{37}$. An alkyl group with 21 carbon atoms would be expected to have $(21 \times 2) + 1 = 43$ hydrogen atoms (not $2n + 2$ because this is an alkyl group and not an alkane). The alkyl group in the fatty acid has six hydrogen atoms fewer than this and therefore three $\text{C}=\text{C}$ bonds.

The relative molecular mass of the fatty acid is 334.60. This acid has 3 $\text{C}=\text{C}$ bonds, and so 3 mol I_2 will react with one mole of it. This means that 334.60 g of the acid would react with 253.80×3 , or 761.40 g of I_2 .

100 g of the acid would therefore react with:

$$\left(\frac{761.40}{334.60}\right) \times 100 = 227.56 \text{ g of } \text{I}_2$$

Therefore the iodine number is 228.

B.3 The relative molecular mass of an oil containing just one type of triglyceride is 909.43. The iodine number of this oil is 112. Deduce the number of $\text{C}=\text{C}$ bonds in each oil molecule.

112 g of iodine reacts with 100 g of oil.

So $\frac{909.43}{100} \times 112 = 1019$ g of iodine react with 1 mol of oil.

1019 g of iodine is $\frac{1019}{253.80} = 4.01$ mol

So 1 mol oil reacts with 4 mol iodine. This means that there are four $\text{C}=\text{C}$ bonds per molecule.

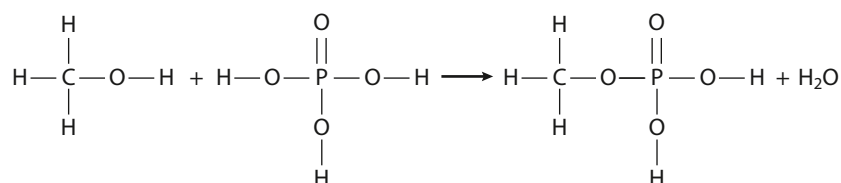
? Test yourself

- 11 Arrange the fatty acids below in order of increasing melting point (lowest first):
 $C_{19}H_{39}COOH$ $C_{17}H_{35}COOH$
 $C_{21}H_{43}COOH$ $C_{17}H_{33}COOH$
- 12 Classify each of the following fatty acids as saturated, monounsaturated or polyunsaturated.
 - a $C_{17}H_{35}COOH$
 - b $C_{21}H_{35}COOH$
 - c $C_{19}H_{37}COOH$
- 13 Calculate the iodine number of each of the fats/oils below. (M_r for $I_2 = 253.80$)
 - a 25 g of oil reacts with 45 g of iodine
 - b 12.2 g of fat reacts with 2.44 g of iodine
 - c 1.53 g of oil reacts with 1.53 g of iodine
- 14 The formulas of some fatty acids are given below:
 - a $C_{13}H_{27}COOH$
 - b $C_{15}H_{23}COOH$
 - c $C_{17}H_{33}COOH$
 - d $C_{19}H_{35}COOH$
 Deduce the number of $C=C$ bonds in each and work out the iodine number.
- 15 A fatty acid with $M_r = 254.46$ has an iodine number of 100. Work out the number of $C=C$ bonds in the fatty acid.
- 16 The relative molecular mass of an oil containing just one type of triglyceride is 885.4. The iodine number of this oil is 57. Deduce the number of $C=C$ bonds in each molecule.

Phospholipids

The second major class of lipid in the body is the **phospholipids**. These are the main lipids found in cell membranes. A membrane made up of a bilayer of phospholipids surrounds each cell and phospholipid membranes also surround many of the inner structures (called organelles) within the cells.

Phospholipids contain phosphate ester groups and the general equation for the formation of a phosphate ester from an alcohol in the laboratory can be represented as:



The reaction does not actually occur like this and polyphosphoric acid is used instead of phosphoric acid.

Phospholipids are similar in structure to triglycerides in that they contain two fatty acid chains esterified (ester linkage) to two of the hydroxyl groups of glycerol (Figure B.28). However, they also contain a phosphate group that has reacted with the third hydroxyl group of glycerol (phosphate-ester linkage).

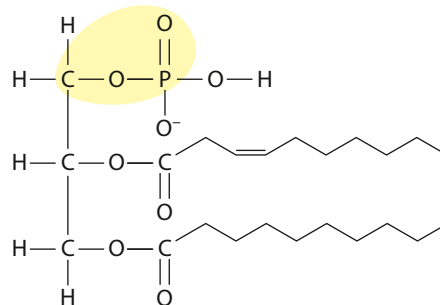


Figure B.28 A phosphate ester group.

The phosphate usually undergoes a condensation reaction with another alcohol to form a second phosphate–ester linkage (Figure B.29) and therefore a **phosphodiester group**.

If the second alcohol is choline (Figure B.30) a **phosphatidyl choline** (also known as a **lecithin**) is formed.

Therefore phospholipids have a polar head (the phosphate ester) and a non-polar tail (the two fatty acids). This property allows them to form bilayers when in an aqueous environment because they arrange themselves so that the polar head groups are in contact with water (forming hydrogen bonds) while the non-polar tails minimise their exposure to the water and interact with each other in the interior of the bilayer (Figure B.31).

There are also phospholipids that are not based on glycerol but rather on sphingosine.

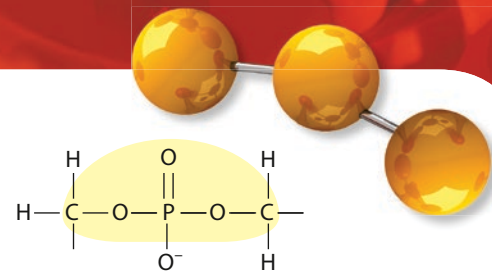


Figure B.29 A phosphodiester group.

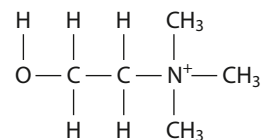


Figure B.30 Choline is an amino alcohol.

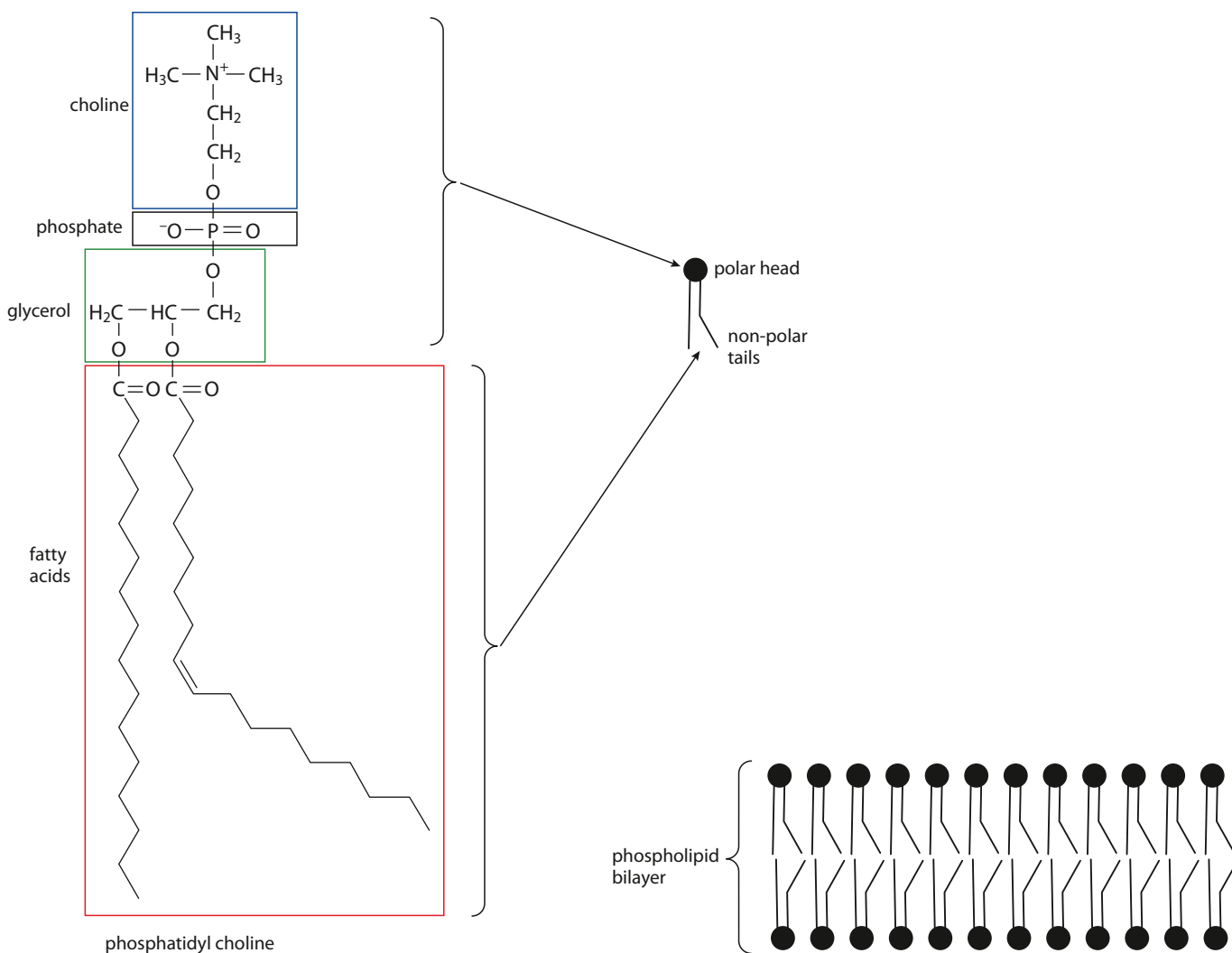


Figure B.31 The structure of phosphatidyl choline and a representation of the lipid bilayer. The fatty acid attached to the central carbon of glycerol is always unsaturated, hence the kink in the chain.

The release of fatty acids when fats undergo hydrolysis is one reason that fats can develop an unpleasant smell when they go rancid.

Hydrolysis of fats

Hydrolysis can occur when a fat is heated with an acid or an alkali. In this reaction, the ester bonds between the fatty acids and the glycerol backbone are broken. This will yield glycerol (propane-1,2,3-triol) and the original fatty acids (Figure B.32).

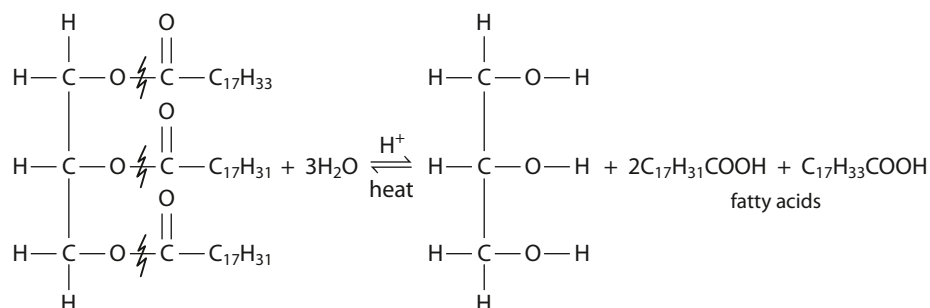


Figure B.32 Hydrolysis of a fat in acidic solution to yield one molecule of glycerol and three molecules of fatty acid.

When the fat is hydrolysed under alkaline conditions, by heating with sodium hydroxide, the sodium salts of the fatty acids are formed; these can be used as soap (Figure B.33).

When fats are ingested, they must be broken down into smaller molecules in order to be absorbed into the body from the intestine. This breakdown is catalysed by enzymes (e.g. pancreatic lipase), which are secreted into the intestine. These enzymes catalyse the hydrolysis of two of the ester bonds in the triglycerides to produce two free fatty acid molecules and a 2-monoglyceride (still one ester linkage on the middle carbon of glycerol), which are then absorbed. Once absorbed, the fatty acids and 2-monoglyceride are reassembled into triglycerides and transported to their destination.

The hydrolysis of phospholipids is more complicated and depends on the reaction conditions. If the phospholipid is heated with a concentrated acid it can be broken down into all its components (Figure B.34) but milder conditions result in only the breaking of the $\text{O}-\text{C}=\text{O}$ ester linkages to yield two fatty acid molecules and a phosphodiester (Figure B.35).

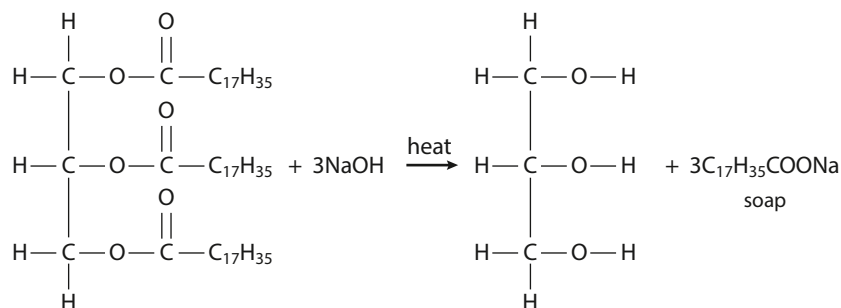


Figure B.33 Hydrolysis of a fat under alkaline conditions.

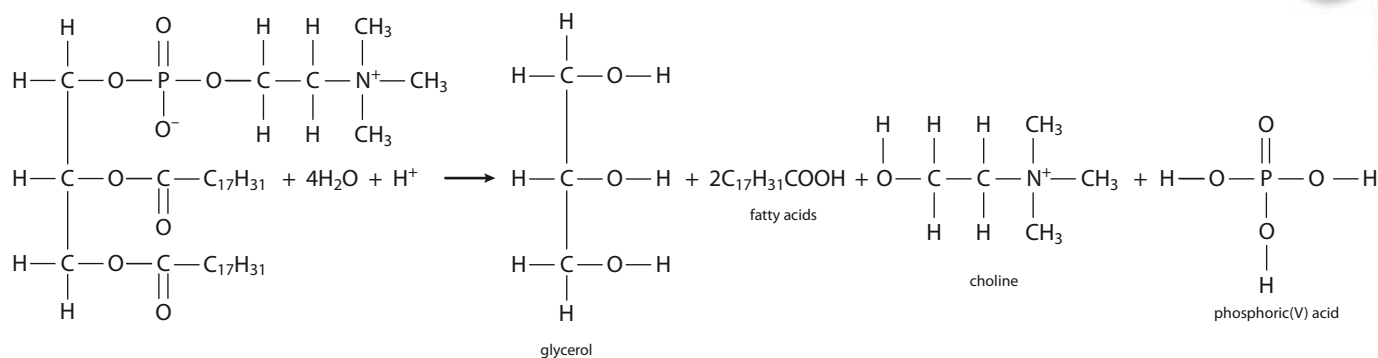


Figure B.34 Hydrolysis of a phospholipid.

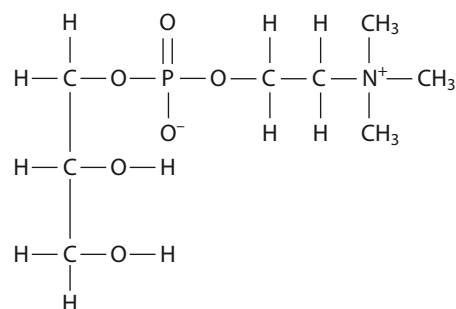
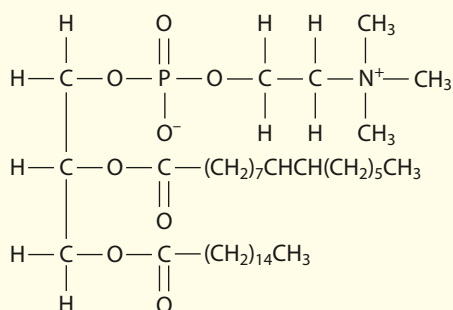


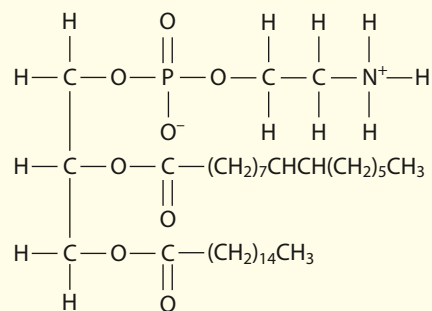
Figure B.35 A phosphodiester formed by mild hydrolysis of a phospholipid.

? Test yourself

- 17** Give the structural formulas of all the products of complete hydrolysis of the phospholipid below using a concentrated acidic solution.



- 18** When the phospholipid below undergoes complete hydrolysis, one product is different from the products in question 17. Give the structural formula of the product that is different.



- 19** Give the structures of the products of the complete hydrolysis of the compound in question 18 under strongly alkaline conditions.

Rancidity

Rancidity refers to the unpleasant odours and flavours that develop when fats go 'bad'. There are two main processes involved – hydrolytic rancidity and oxidative rancidity.

Hydrolytic rancidity

This is the process by which fats and oils are broken down into their constituent fatty acids and glycerol (refer back to Figure B.32).

These fatty acids can have unpleasant smells – butanoic acid smells of vomit. The ester bonds between the fatty acid and the glycerol are broken. The process can be speeded up by increased water content, the presence of enzymes (e.g. lipase), high pH values (increased concentration of OH^-) or low pH values (increased concentration of H^+) and raised temperatures. During deep-fat frying, the extremely high temperatures and presence of water in foods can lead to rapid rancidification.

Oxidative rancidity

Fatty acid chains of lipids can undergo oxidation in a free radical chain reaction; enzymes may also catalyse the mechanism. In this process, oxygen adds across the $\text{C}=\text{C}$ double bonds of unsaturated fatty acids and causes them to break down into unpleasant-smelling and -tasting, volatile compounds such as ketones and aldehydes as well as some alcohols. The greater the degree of unsaturation, the more predisposed a lipid will be to oxidative rancidification.

Other factors that speed up the rate of oxidation include increased light intensity, increased temperature, high availability of oxygen and the amount of metals such as copper present. The rate of oxidation is also influenced by water content.

Steroids

The third main class of lipid in the body is the **steroids**. Examples include cholesterol, the sex steroids (testosterone and oestrogen) and adrenal hormones (hydrocortisone and aldosterone).

Steroids are hydrophobic (mostly non-polar) molecules having a common structure, known as the **steroid backbone** (Figure B.36). This is made up of three six-membered rings (called A, B and C) and a five-membered ring (called D) fused together. The steroids vary depending on the type and position of substituents on the steroid backbone. There is also usually a carbon-carbon double bond in either ring A or ring B. Oestrogens are different to the other steroids in that they have an aromatic A ring (benzene ring).

Cholesterol (Figure B.36) is a major steroid found in the body. It is found in cell membranes, where it maintains fluidity of the membrane, and it is also the precursor of other steroids such as those mentioned above, as well as bile acids and vitamin D. Although cholesterol plays an important role in the body, it can also have a negative effect on health in that it can contribute to heart disease.

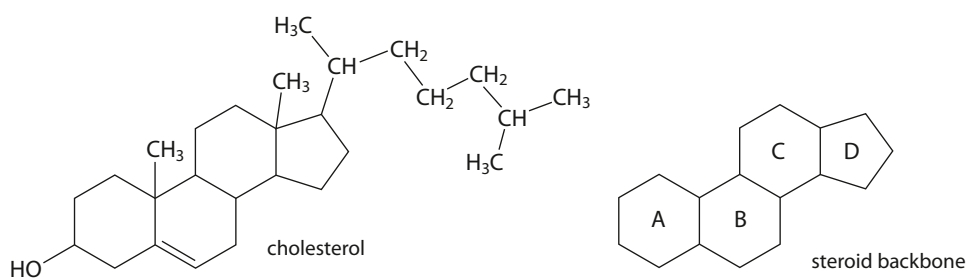
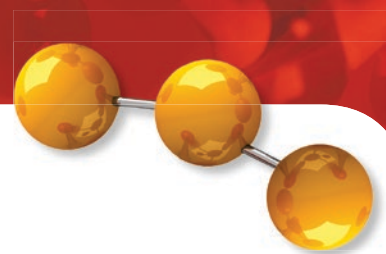


Figure B.36 Structure of cholesterol and the steroid backbone.



The use and abuse of steroids

As well as the uses of cholesterol described above, steroids have an important use in the body as hormones. Hormones are chemical messengers – they allow cells in the body to communicate. They cause a specific effect on hormone-sensitive cells (called target cells) in the body – the effect they cause depends on the type of hormone and the type of target cell. The steroid hormones include the corticosteroids (e.g. aldosterone) and the sex steroids (e.g. oestrogen and testosterone).

Steroids are used medically as oral contraceptives. They can also be used for hormone-replacement therapy (HRT) in women who are going through the menopause. During the menopause, the ovaries stop producing oestrogen and so the level of oestrogen in the body drops dramatically. This leads to symptoms such as night sweats, sleeplessness and hot flushes. HRT replaces the hormones to alleviate the symptoms of menopause. Because HRT contains oestrogen, its long-term use has been linked with an increase in the risk of breast cancer, because some types of breast cancers are dependent on oestrogen for their growth. Once HRT is stopped, however, a woman's risk of breast cancer starts to return to that of the general population within three years of stopping treatment.

Other clinical uses of steroids include the treatment of inflammation and associated conditions. Hydrocortisone (also known as cortisol) is a naturally occurring corticosteroid, so-called because it is produced in the adrenal cortex in the adrenal glands (there are two adrenal glands in the body – one above each kidney). One of its effects is to reduce inflammation and therefore it, along with synthetic corticosteroids, is used clinically for a variety of conditions such as eczema, rheumatoid arthritis and asthma.

Testosterone is an androgen (male sex hormone). It has medical uses as an androgen replacement in men who have testosterone deficiency – for example, because of a disorder of the testes. Androgens are also used non-medically – for example, in sport.

Androgen abuse involves their use for non-medical purposes – they are called **anabolic steroids** because they promote tissue growth, in particular of muscle. The three most common anabolic steroids that are abused are testosterone and the synthetic derivatives nandrolone and stanozolol. They are taken by sportsmen and sportswomen to enhance performance because they increase muscle mass and are also believed to improve endurance. They have been used in disciplines such as athletics, weightlifting and cycling. The ethical implications for taking anabolic steroids are clear, in that it gives that person an unfair advantage over their competitors and is simply cheating. It is a major concern for sporting bodies worldwide, and random drug screening in major sporting events is routinely employed to detect abuse. In the wider community, anabolic steroids are also used for bodybuilding and by a minority of people in certain occupations, such as security guards and 'bouncers', for cosmetic reasons to give themselves a more masculine and intimidating look.

Abusing anabolic steroids can have a major impact on the body – their use can cause a number of side effects, such as breast growth in men, acne, infertility, mood swings and aggressiveness. They can also cause high blood pressure, liver disease (including cancer), heart attack and stroke.

Psychologically, abusers of anabolic steroids can become addicted to them, developing an increased desire to keep taking them even if unwanted side effects occur.

The effects of lipids on health

Transport of lipids

Triglycerides and cholesterol are essentially insoluble in aqueous blood plasma and therefore cannot be transported as free molecules. Instead these lipids assemble with phospholipids and proteins to form particles known as **lipoproteins**. The inside of a lipoprotein is hydrophobic, but the outside is hydrophilic and, therefore, can travel in the bloodstream. There are different types of lipoprotein which are classified according to their relative densities. Each type has a different composition of protein, triglycerides and cholesterol.

Low-density lipoproteins (LDLs) consist mainly of cholesterol and are the major reservoir of cholesterol – they transport it throughout the body, where it can be stored in the tissues or used (for example, in cell membranes). High levels of LDL-cholesterol can lead to fatty deposits in the walls of arteries in the heart and elsewhere, leading to atherosclerosis (hardening of the arteries) which can result in heart attack and stroke.

High-density lipoproteins (HDLs) are protein-rich particles. They are smaller than LDLs and denser because they contain more protein (protein is more dense than lipid). HDLs consist of approximately 33% protein, compared with 25% for LDLs. HDLs scavenge cholesterol from LDLs, tissues and artery walls and return it to the liver, where it is converted to bile acids. HDLs remove cholesterol from the tissues and arteries and have a beneficial effect with regards to heart disease.

LDL-cholesterol levels are not based on the amount of cholesterol from the diet but rather on the amount and type of fat taken in. Saturated fats (especially myristic, palmitic and lauric acids) and *trans* fats increase the level of LDL-cholesterol in the body and so are associated with an increased risk of heart disease. *Trans* fats have also been found to lower HDL-cholesterol levels.

Polyunsaturated and monounsaturated fats, however, have been shown by some studies to lower LDL-cholesterol and to increase HDL-cholesterol. Therefore the majority of fat in the diet should be in the form of mono- and polyunsaturated fats, while intake of saturated fats and *trans* fats should be limited (or totally excluded) in order to reduce LDL-cholesterol levels.

Nature of science

Taking drugs to enhance performance in sport is as old as sport itself. What has changed, however, is the ability of scientists to detect drugs. Advances in technology and greater understanding of metabolic pathways have allowed the development of ever more sophisticated testing protocols for drugs and their metabolites in sport. But science is always improving and so blood/urine samples from the Olympics are kept for eight years so that future advances might be able to detect previously undetectable substances.

There have been a great number of studies in the past few years looking at the connections between fat intake, especially saturated fats, and cardiovascular disease. However, the results from these studies are far from

Low LDL-cholesterol levels and high HDL-cholesterol levels reduce the risk of heart disease.



conclusive in establishing a causal link and recent studies have suggested that saturated fats are not as bad as once thought. The prevailing belief among the general public is that saturated fats cause heart disease and this has led to the development of many low-fat food products. These low-fat products, however, can often be unhealthy in other ways, for example having higher sugar content. The role of the scientist is essential in improving public health but it is not just a case of carrying out experiments and trials but also in informing the public in a clear and unbiased way so that they are able to make their own decisions. The debate about the health benefits of saturated fats is likely to go on and on and it will be important to have clear information that avoids over-simplification of the issues.

B4 Carbohydrates

Carbohydrates are widespread in nature and are, in fact, the most abundant class of biological molecules. They range from simple sugars such as glucose and fructose, to more complex carbohydrates such as starch and cellulose. The simplest of the carbohydrates are called **monosaccharides** – these can exist on their own or can join together to form polymers known as **polysaccharides**.

Carbohydrates have a number of functions in the human body including:

- an **energy source** – glucose is converted into ATP to drive biochemical reactions
- an **energy store** – glucose is polymerised into glycogen and stored for those times when blood glucose levels fall and energy is needed.

ATP is the energy currency of cells. At a simple level, it is a molecule made in endothermic processes and the energy released when it is hydrolysed later can be used to make unfavourable reactions work.

The formulas of carbohydrates

Carbohydrates have the general formula $C_x(H_2O)_y$.

Glucose and sucrose (common table sugar) are examples of carbohydrates. Glucose has the molecular formula $C_6H_{12}O_6$, which can be written as $C_6(H_2O)_6$, whereas sucrose has the molecular formula $C_{12}H_{22}O_{11}$, which can be written as $C_{12}(H_2O)_{11}$.

Monosaccharides

The smallest monosaccharide molecules contain just three carbons and are known as **trioses**. We will, however, concentrate on those monosaccharide molecules that contain six carbons – the **hexoses**. Examples of hexoses are glucose and fructose – both have the same molecular formula ($C_6H_{12}O_6$) but they have different structural formulas. They can exist in either the straight-chain form or the ring form (Figure B.37).

Glucose contains an aldehyde group and is known as an **aldose** sugar, whereas fructose contains a ketone group and is known as a **ketose** sugar.

Learning objectives

- Describe the main functions of carbohydrates in the human body
- Understand that carbohydrates have the general formula $C_x(H_2O)_y$
- Describe the common structural features of monosaccharides
- Understand that monosaccharides can exist in straight-chain and ring forms
- Use Haworth projections to represent cyclic monosaccharides
- Draw the structures of disaccharides and polysaccharides that can be formed in condensation reactions of monosaccharides
- Understand that glycosidic bonds are formed between monosaccharides when they form disaccharides and polysaccharides

Monosaccharides have the empirical formula CH_2O . They contain a carbonyl group (a ketone or aldehyde) and have at least two hydroxyl (OH) groups as part of their structure.

The ring form of monosaccharides

In solution, the straight-chain form of a monosaccharide cyclises into a ring structure. This happens when the carbon with the double-bonded oxygen (C1 in glucose/C2 in fructose) joins to C5 via a bridging oxygen atom (Figure B.37).

This forms an ether linkage, C–O–C, within the ring (Figure B.38).

The carbon atom that was originally part of the carbonyl (aldehyde/ketone) group in the straight chain and is now part of the ether linkage is sometimes called the anomeric carbon.

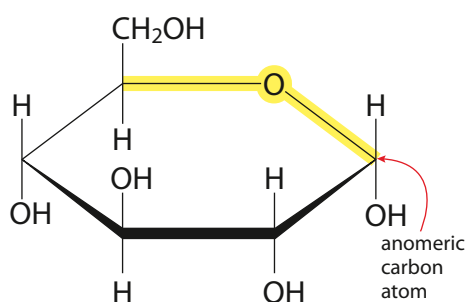


Figure B.38 The ether linkage forming a ring monosaccharide.

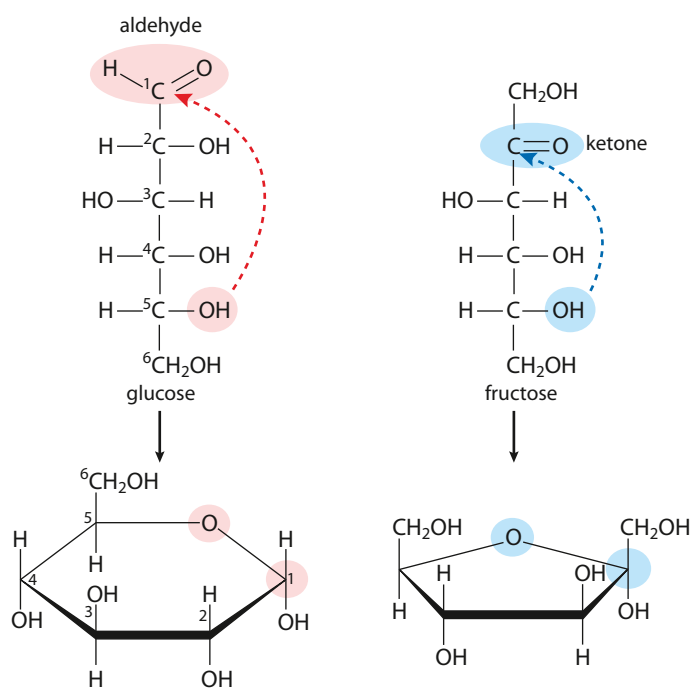
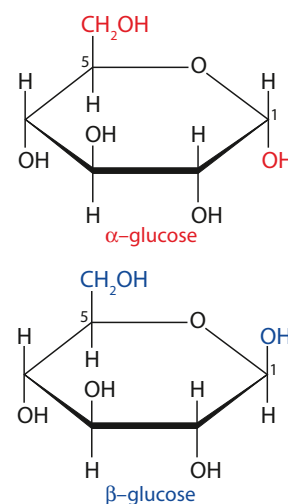


Figure B.37 Straight-chain and ring structures of glucose and fructose.

Extension

The cyclisation reaction can happen in two ways, so that when an OH group is formed on C1 it can either be on the same side of the ring as the CH_2OH group on C5 (β -isomer) or on the opposite side of the ring (α -isomer). α - and β -forms of glucose are called anomers. α - and β -glucose are discussed more fully in the Higher Level section later.





The ring forms shown here are called **Haworth projections** and are a convenient way of representing the ring form of a monosaccharide. The carbon atoms are not shown but there is one at each vertex in the ring. The hydrogen atoms directly attached to carbon atoms in the ring are also sometimes omitted. The ring is shown as planar and perpendicular to the plane of the paper with the thicker bonds representing C–C bonds closer to us. The oxygen of the ether link is shown at the back right for six-membered rings and back centre for five-membered rings.

Disaccharides

When two monosaccharides join together, they form a **disaccharide**.

Examples of common disaccharides are:

- **maltose** – one of the products from the digestion of starch; made up of glucose + glucose
- **lactose** – a major sugar found in milk; made up of glucose + galactose
- **sucrose** – common table sugar; made up of glucose + fructose.

Formation of a disaccharide is another example of a condensation reaction, when two molecules join together with the elimination of water. This results in the formation of a **glycosidic bond** (linkage) between the two sugars (Figure B.39).

A glycosidic bond is an ether group and forms from the anomeric carbon on one ring to an OH group on another ring. The linkage shown in Figure B.39 is from C1 to C4 and so is a 1,4-glycosidic bond, but 1,2-glycosidic bonds, 1,3-glycosidic bonds etc. also form between monosaccharides. In order to predict the disaccharide formed from two monosaccharides, you need to know how they will join together – different disaccharides are formed depending on how the molecules join together.

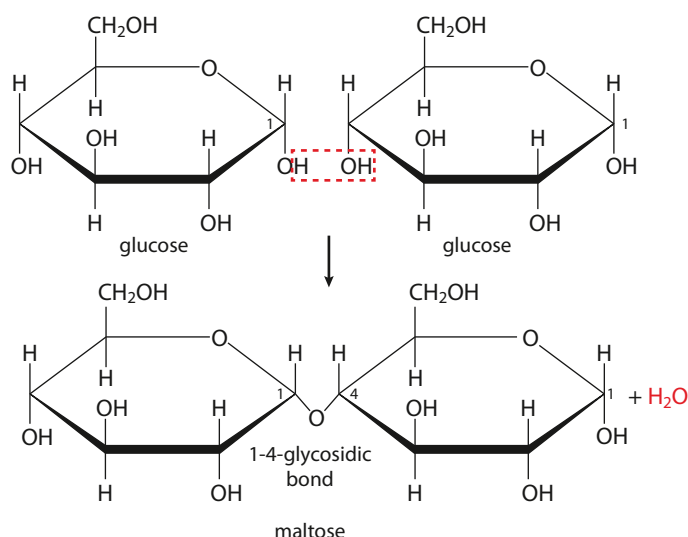


Figure B.39 The condensation reaction between two glucose monosaccharides to produce maltose, a disaccharide.

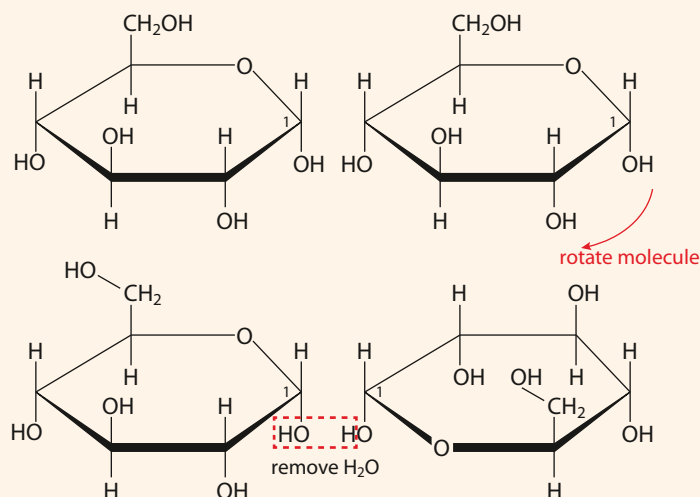
Extension

The rings in monosaccharides are not planar.

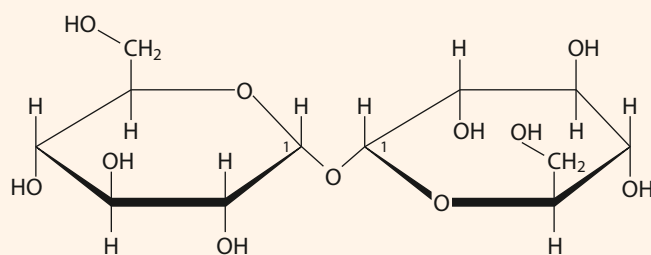
Worked example

B.4 Draw the disaccharide formed when two glucose molecules form a 1,1-glycosidic bond.

Two glucose molecules must be drawn, but one is rotated around so that the C1 atoms are next to each other.



Once the OH groups on each C1 are next to each other, water is removed and a glycosidic bond formed:



Monosaccharides can also form glycosidic bonds in different ways – for example an alternative 1,4-linkage is formed in Figure B.40.

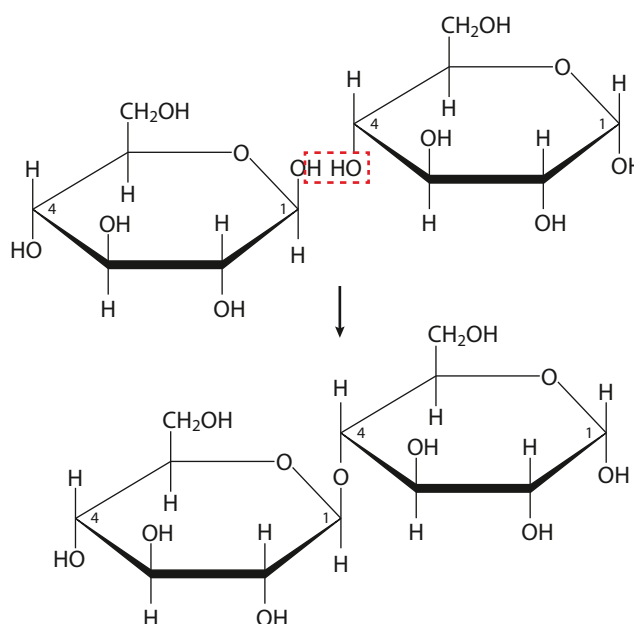
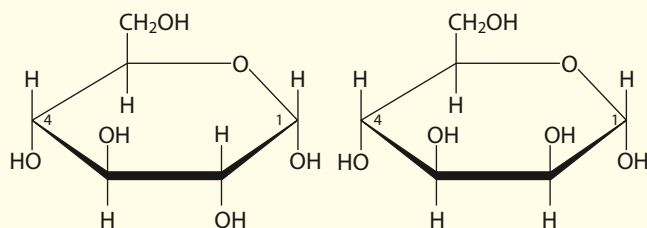


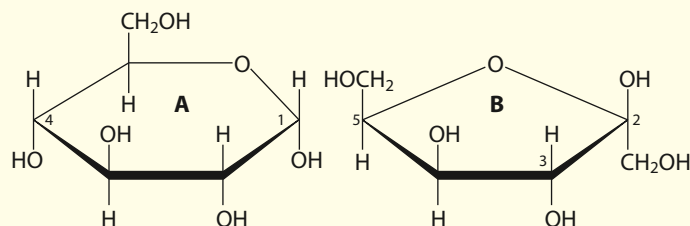
Figure B.40 The formation of a different type of 1,4-glycosidic bond.

? Test yourself

20 Draw the two possible disaccharides formed with a 1,4-linkage between these two monosaccharides:



21 Draw the disaccharide formed when monosaccharide **A** joins to monosaccharide **B** with a 1,3-glycosidic bond.



Polysaccharides

Polymers of many monosaccharide molecules joined together by condensation reactions are known as **polysaccharides**. Examples include **starch**, **glycogen** and **cellulose**, which are all polymers of glucose. For instance, plants convert excess glucose into starch for storage, and starch consists of a mixture of two types of glucose polymers, called **amylose** and **amylopectin**. Amylose consists of thousands of glucose units linked together to form linear, unbranched chains (Figure B.41), whereas amylopectin is a branched polysaccharide polymer in which glucose units are linked by 1,4-glycosidic bonds and 1,6-glycosidic bonds at the branch points.

Humans have digestive enzymes that can hydrolyse the glycosidic bonds in starch – hence we can break down starch polymers into smaller fragments and eventually into single glucose units which can then be absorbed into the bloodstream and used as an energy source.

Just as plants store their excess glucose in the form of starch, we store our excess glucose in the form of glycogen. Glycogen is a branched polymer of glucose units linked by 1,4-glycosidic linkages and 1,6-glycosidic linkages. Glycogen is stored as granules within cells (in particular, liver and skeletal muscle cells) and is converted back into glucose when energy is needed.



Trade in sugar was one of the major driving forces behind the development of the slave trade between west Africa and the Caribbean in the 17th to 19th centuries. Millions of Africans were enslaved and taken to work on sugar plantations to make money for wealthy British owners and to fund the expansion of the British Empire.

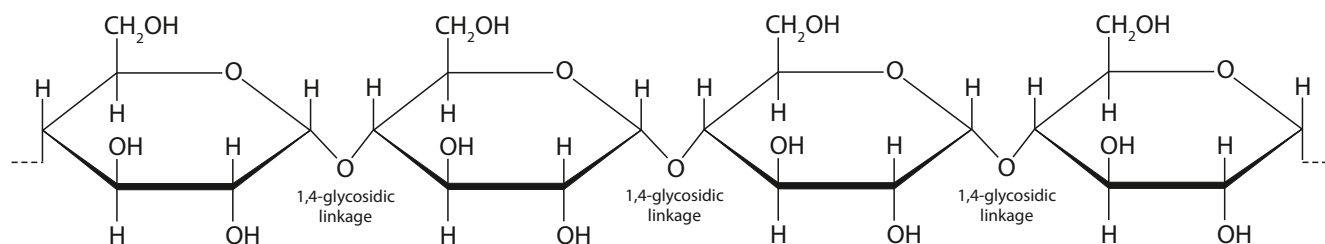


Figure B.41 Glucose units are linked by 1,4-glycosidic bonds in the polysaccharide amylose.

The relationship between properties, function and structures

Monosaccharides contain several OH groups which are able to form hydrogen bonds to water molecules – therefore they are soluble in water. Glucose is an important source of energy in the body (it is the main fuel for the brain and a key energy source for muscles) and because it is soluble in water it can be transported easily around the body in the blood to where it is needed.

In order to maintain glucose levels in the blood at the correct level for normal brain functioning (for example), stores of glucose must be maintained in the body. Because glucose is soluble in water, if large amounts were stored in cells then it would be dissolved in the cytoplasm and affect osmotic pressure. The concentration of the solution in the cells would be high and large amounts of water would flow into them by osmosis and cause the cells to burst. This is why glucose is converted into glycogen which has very large molecules and, even though it is also soluble in water (due to large number of oxygen atoms in the molecules and OH groups that can participate in hydrogen bonding), it is not as soluble as glucose. The large polymeric structure of glycogen allows it to be stored in a different form – as granules in association with proteins – which has little effect on osmotic pressure. Glycogen can be readily broken down to glucose when it is needed.

Cellulose is another polysaccharide made up of linear chains of glucose molecules joined together. The polymer chains can pack closely together and many hydrogen bonds are formed between the chains, giving the cellulose structure significant strength. Because cellulose is insoluble in water and is also strong, it is an important structural polysaccharide in plants. The properties of cellulose are very different from those of starch and glycogen because the glucose units are joined together in a different way in starch and glycogen, which means that hydrogen bonding between chains is not as extensive as in cellulose. Also the presence of branches in the polymer chains in starch/glycogen means that they cannot pack together as closely and therefore there are weaker forces between them.

Nature of science

Diabetes is a disease caused by the inability of the body to control the level of blood sugar. Although the disease has been known since ancient times, treatments for the disease have only been available for about 100 years. Observation is extremely important in science and recognition of the symptoms of a disease is the first step towards discovering a cure. Each generation of scientists stands on the shoulders of previous generations (or as Isaac Newton put it: 'If I have seen further it is by standing on the shoulders of giants') and pioneering work in the 19th century and a great deal of inspiration and hard work led to the discovery of the link between diabetes and insulin, the isolation of insulin and the development of an effective treatment for diabetes.

Simplified structures are often used in science to aid in visualising complex molecules; the Haworth projection is one such structure.

B5 Vitamins

Vitamins are micronutrients

Nutrients are chemical substances derived from food that are used by the body for growth and survival. So far we have discussed proteins, carbohydrates and fats. All these are required in relatively large amounts in the diet – they are known as **macronutrients**.

Vitamins are organic molecules that are classified as **micronutrients**. These are essential in small amounts (<0.005% body weight) for the normal functioning of the body. They are required in mg or µg amounts.

Vitamins cannot generally be made in the body (exceptions include vitamin D) so they must be obtained from the ingestion of suitable foodstuffs as part of a healthy diet. They are also sometimes taken in the form of food supplements – vitamin tablets.

Classification of vitamins

There are two main classes of vitamins – **water-soluble** vitamins and **fat-soluble** vitamins. Water-soluble vitamins have molecules with structures containing many polar groups – for example, OH groups, which can hydrogen bond to water molecules. Examples include vitamin C and the B group of vitamins (consisting of eight different vitamins) – these are excreted readily in the urine and stores are depleted rapidly, so a daily intake is required.

Fat-soluble vitamins are mostly non-polar in nature, with molecules containing long hydrocarbon chains or rings. Examples include vitamins A, D, E and K. This type of vitamin is stored in the body and therefore, if excessive amounts are taken, levels can build up resulting in toxicity.

Fat-soluble vitamins are not soluble in water and water-soluble vitamins are not soluble in fat.

Vitamin C

This is a water-soluble vitamin because its molecules contain several OH groups (and other oxygen atoms) that enable them to form hydrogen bonds with water (Figure B.42). Vitamin C (ascorbic acid) plays a key role in tissue growth and repair – more specifically it is required for the synthesis of collagen, a protein found in connective tissue (for example, in bone, skin and blood vessels). It can also act as an antioxidant, protecting the body from damage by free radicals produced naturally during normal metabolic processes.

Vitamin C is found widely in fresh fruit and vegetables, particularly in citrus fruits. Deficiency of vitamin C results in the disease called scurvy – symptoms of which include swollen, bleeding gums, muscle and joint pain, and poor healing of wounds.

Vitamin A

Vitamin A (retinol) is a fat-soluble vitamin (Figure B.43). It contains a long hydrocarbon chain and a hydrocarbon ring and, although it contains an OH group, the polar nature of this group is not enough to offset the non-polar nature of the rest of the molecule. Having a mostly non-polar structure means that vitamin A is soluble in fat rather than water.

Learning objectives

- Understand what vitamins are
- Compare the structures of vitamins A, C and D and relate their structures to their relative solubilities in water or fat
- Understand that some vitamins are sensitive to heat
- Discuss the different vitamin deficiencies around the world the diseases to which they lead
- Suggest ways of solving deficiency problems

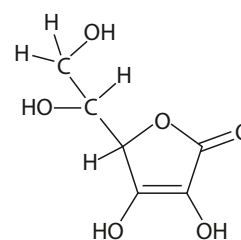


Figure B.42 Structure of ascorbic acid (vitamin C), a water-soluble vitamin.

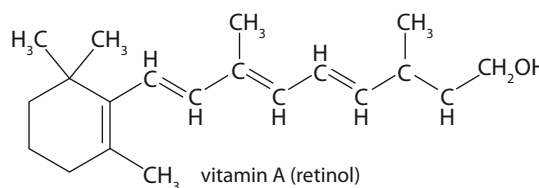


Figure B.43 Structure of the fat-soluble vitamin A.

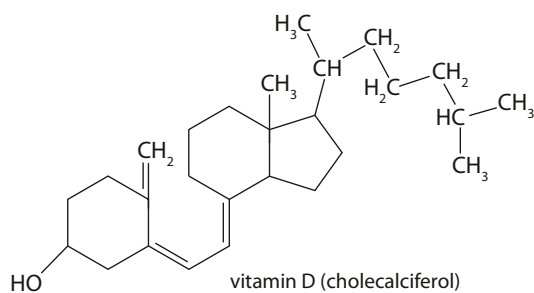


Figure B.44 The fat-soluble vitamin D.

Exam tip

You should be able to work out from its structure whether a particular vitamin is water- or fat-soluble. Vitamins containing many OH groups and/or several very electronegative atoms (such as nitrogen or oxygen) are generally water-soluble, and those that consist almost entirely of carbon and hydrogen are fat-soluble.

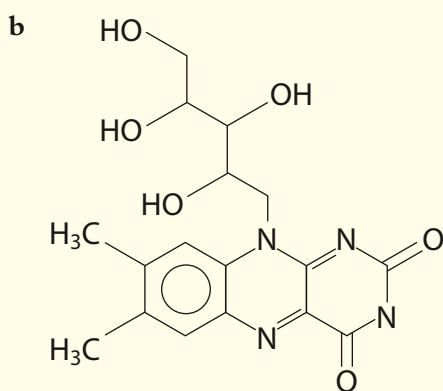
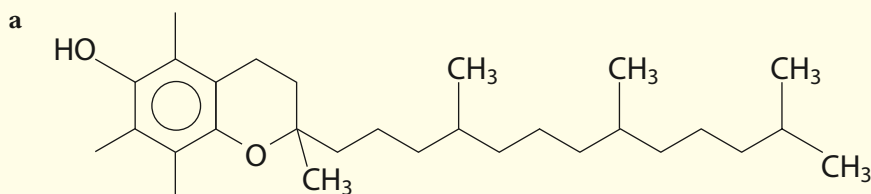
Vitamin A is important for vision, especially in low light intensities. It also plays a role in growth and development, in skin repair and in the immune system. Vitamin A (in the form of retinol) is found in animal products such as liver, egg yolk and dairy products. It can also be formed in the body from the provitamin β -carotene, a vitamin A precursor, found widely in fruit and vegetables such as carrots. Deficiency of vitamin A results in a condition known as xerophthalmia, a severe drying of the eye accompanied by night blindness – it is a leading cause of blindness in children in developing countries.

Vitamin D

Vitamin D (cholecalciferol) has the structure shown in Figure B.44. It contains one polar OH group and a large non-polar hydrocarbon backbone, making it predominantly non-polar and therefore fat-soluble. Vitamin D plays an important role in promoting the absorption of calcium and phosphorus from food and in promoting the mineralisation of bone. It is provided by butter, cheese, milk and fish-liver oil. Vitamin D can be synthesised in the body by the action of sunlight on provitamins in the skin, but deficiency can occur in those with limited exposure to sunlight and also in children, who require higher levels for growth. Vitamin D deficiency in children can lead to a condition called rickets, characterised by the softening and deformity of bones.

? Test yourself

22 The structures of two vitamins are given below. Classify each as water-soluble or fat-soluble.





The effect of heat on vitamins

Some vitamins, for example, vitamin C and thiamin (vitamin B₁), are sensitive to heat. This can be important because many foods that we consume are processed (e.g. pasteurisation) in some way and this processing can cause a reduction in the vitamin content.

Cooking can also reduce vitamin content – vitamin C will be lost when vegetables such as broccoli are boiled in water, not only due to the effect of the heat but also because the vitamin C is soluble in water and will dissolve in the cooking water.

Vitamin deficiencies

‘Malnutrition’ is the term used to describe an inadequate intake of the nutrients needed to maintain good health. It can be caused by not eating enough food and also by eating a poorly balanced diet – for example, a diet of processed, fast foods that lack essential vitamins and minerals.

Vitamin deficiencies affect millions of people worldwide and the main cause of this is the lack of availability of enough food and/or the right types of food. The ideal solution to vitamin deficiency would be to make fresh, vitamin-rich food available to everyone. Because this is unlikely to happen, other approaches have been tried.

Vitamin A deficiency

Deficiency of vitamin A can result in a condition called xerophthalmia, which is a leading cause of blindness in many developing countries. Fortification of foods with vitamin A (adding vitamin A to foods) has proved a successful strategy for combating this deficiency. Fortification of margarine has produced great success in many countries and other foods are showing promise – for example, sugar fortification (used in Guatemala) and maize fortification (in Zimbabwe).

Programs also exist to encourage farmers to grow varieties of foods richer in provitamin A (which can be converted into vitamin A in the body) – this is called biofortification. For instance the introduction of orange-fleshed sweet potato into Uganda to replace the indigenous white-fleshed variety has met with some success in reducing vitamin A deficiency.

A more controversial form of biofortification involves genetic modification of foodstuffs to make them richer in a particular vitamin. Genetic modification (GM) has been used to produce ‘golden rice’ – a variety of rice rich in provitamin A. It is hoped that the use of golden rice will make a significant difference to vitamin A deficiency in countries such as India, Bangladesh and Vietnam.

Vitamin B group deficiency

Vitamin B₃ (called **niacin**) is converted to a coenzyme that plays a key role in oxidation–reduction processes in the cell. Deficiency results in a condition called pellagra, characterised by diarrhea, dermatitis and dementia. Vitamin B₁ (**thiamine**) is converted to a coenzyme that is essential for energy production within cells. Deficiency leads to beriberi, characterised by muscle weakness. Many foods, such as breakfast cereals, are fortified with niacin and thiamine, and deficiency is rare in developed countries.

Vitamin C is not ‘denatured’ by the effect of heat but rather undergoes an oxidation reaction to form initially dehydroascorbic acid.



Many foods contain food supplements, such as vitamins, which are introduced to improve health but what are the ethical implications of ‘force feeding’ these to people? To what extent must the need for personal choice be balanced against the greater good – introducing food supplements should make people healthier and reduce the burden on health services but can governments force people to become healthier? If this is acceptable, should alcohol and cigarettes be completely banned?



Some international environmental organisations such as Greenpeace are opposed to the use of GM crops – see page 75 for some of the arguments against the use of genetically modified organisms.

Vitamin C deficiency

Vitamin C deficiency results in scurvy. This used to be a common problem among sailors, who spent long periods at sea without fresh fruit and vegetables – then it was recognised that a regular intake of citrus and other fruits and vegetables would prevent this disease. Nowadays, scurvy is rare in developed countries.

Vitamin D deficiency

Vitamin D deficiency can result in rickets in children. As explained above, it is a condition in which softening and deformity of the bones occur due to a reduction in the uptake of calcium and phosphate from food. Fortification of dairy products with vitamin D means that deficiency is now rare in industrialised countries. However, it is still a problem in some developing countries where intake of dairy products may be low or where religious or social customs (wearing clothes that cover virtually all the body) and/or climatic conditions prevent an adequate exposure to sunlight.

Nature of science

Science is not without controversy and it is definitely not the case that all scientists believe the same things. One of the greatest chemists of the 20th century was the American scientist Linus Pauling (the only person to win two unshared Nobel prizes – one for Chemistry and one for Peace). Later in life he became obsessed with the idea that high doses of vitamin C could prevent a number of diseases, including the common cold, and even be a treatment for cancer. His claims and the scientific studies associated with them were very controversial and the debate still goes on today about the benefits of high doses of vitamin C.

Careful observation is essential in science and the idea of vitamins originated when it was observed that certain diseases arose when people consumed diets deficient in certain foods. For instance, sailors often developed scurvy when they went on long journeys and did not have access to citrus fruit or that people fed on white rice developed beriberi but those fed on brown rice did not.

B6 Biochemistry and the environment

Xenobiotics

Xenobiotics are compounds that are present in living organisms but should not normally be found there.

These compounds are usually produced industrially and include medicines, drugs, pesticides, plasticisers and dyestuffs. The presence of pharmaceuticals, including antibiotics and hormones, in waste water is becoming a big problem. Pharmaceuticals can enter the water supply in various ways:

- excreted in urine and feces
- washing and showering – pharmaceuticals are found in sweat and so will enter waste water when we wash
- disposal of unwanted medicines – for example, by flushing old medicines down the toilet or by throwing away used dermal patches
- agriculture – from drugs given to animals.

These pharmaceuticals are then carried to sewage treatment plants. However, because of the diverse nature of the chemicals involved, they are not removed from the waste water effectively and some are released into the environment again – this is also a relatively new problem and sewage treatment plants were not designed to remove them. The result of this is that water containing a variety of pharmaceuticals is used for drinking and the irrigation of agricultural land. Although these pharmaceuticals are found in drinking water in only very small amounts (typically ng dm^{-3}), there are concerns that long-term exposure could result in damage to human health.

The release of antibiotics into the environment in waste water is regarded as a particular problem because not only can they cause damage to aquatic organisms, but they can also result in increased resistance of bacteria to antibiotics. Antibiotics (antibacterials) are used to treat a variety of conditions but if bacteria develop resistance to antibiotics such as penicillin these diseases can become much more difficult to cure.

Biodegradable plastics

Plastics derived from alkenes – such as poly(ethene), poly(propene) and poly(chloroethene) – are non-biodegradable, which means that they cannot be broken down by microorganisms when, for instance, they are buried in soil. They are non-biodegradable because of the strong carbon–carbon covalent bonds in the polymer chains. This means that plastics are difficult to dispose of and the three main methods for dealing with waste plastic are burying in a landfill site, incineration and recycling. One solution to the disposal problem is to develop biodegradable/compostable plastics.

Starch has been used widely in the development of bioplastics (plastics from renewable materials) and biodegradable/compostable plastics. Examples of starch-based bioplastics include thermoplastic starch and polylactic acid (PLA).

Learning objectives

- Understand what is meant by xenobiotics
- Understand what is meant by biodegradable/compostable plastics
- Understand the role of starch in the manufacture of biodegradable/compostable plastics
- Understand what is meant by host–guest chemistry
- Discuss an example of the use of host–guest chemistry in the removal of harmful substances from the environment
- Understand some of the uses of enzymes
- Understand what is meant by biomagnification
- Discuss an example of biomagnification
- Understand how the principles of green chemistry could be applied to biochemistry

Large pharmaceutical molecules can be removed very effectively from waste water by the process of reverse osmosis, but this is expensive.



Bisphenol A (BPA), a substance present in some food packaging and drink bottles, has been suggested as one of the causes of low sperm count in men. Its use has been controlled in some countries.



Some European countries, such as Denmark and Switzerland, incinerate large proportions of their waste – other countries, such as Spain, Finland and Ireland, predominantly use landfill sites.

'Biodegradable' and 'compostable' are not the same thing – 'biodegradable' just refers to the fact that the plastic will be broken down by microorganisms such as bacteria. In order for a plastic to be 'compostable' it must be broken down by microorganisms at a rate comparable with that of naturally occurring polymers, such as cellulose, and not produce any toxic products.

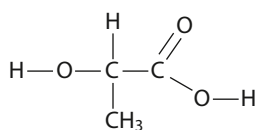


Figure B.45 Lactic acid.

Thermoplastic starch is obtained by mixing starch with plasticisers such as water, glycerol and sorbitol. The plastic obtained does not have very good mechanical and physical properties and therefore it is usually blended with other polymers – biodegradable or non-biodegradable. When blended with biodegradable polymers it can produce polymers that are fully biodegradable, but when blended with non-biodegradable ones only the starch portion will biodegrade.

Because starch is an important energy-storage material found in plants, enzymes are present in organisms to break it down to glucose, which can be broken down further, in cellular respiration, to carbon dioxide and water. So, starch is readily broken down in the environment.

Poly(lactic acid) (PLA) is a polyester derived from lactic acid (2-hydroxypropanoic acid) (Figure B.45). Lactic acid can be obtained from corn starch by fermentation using microorganisms. The plastic formed is biodegradable under certain conditions because of the ester groups between the monomers. It has found uses as a packaging material, plastic cups etc.

However, there is a debate about how environmentally friendly PLA and similar plastics are. Objections to its use include the fact that vast areas of land are given over to growing corn to make into plastics, rather than food. Also the corn is a GM crop, and PLA will only degrade at a measurable rate in an industrial composter.

Host–guest chemistry

An example of host–guest chemistry is the interaction between an enzyme and a substrate. The substrate (guest) does not form covalent bonds to the groups in the active site of the enzyme (host), rather it is held in place by other types of interactions such as hydrogen bonds, London forces, ionic interactions etc. The type of complex formed between an enzyme and a substrate is the basis of a branch of chemistry called 'host–guest chemistry'.

Enzymes can 'recognise' particular substrate molecules and it is this type of molecular recognition that is an important goal of host–guest chemistry. Host molecules can be synthesised that are not only selective but have also been shown to catalyse reactions of guest molecules, so mimicking the role of enzymes.

There are many types of host–guest complexes that have applications in the environment with regard to removing toxic materials. One class that has been used for removal of heavy metal ions from solutions is the calixarenes. Calixarenes have the basic structure shown in Figure B.46a and form cup-like molecules – Figure B.46b. Calixarenes have been used to remove highly radioactive caesium ions from radioactive wastes and to extract uranium ions from water. The cup-like shape (similar to the active site of an enzyme) makes it size-selective and ion–dipole interactions can form between the caesium ions and the oxygen atoms of the OH groups.

Host–guest chemistry is a form of supramolecular chemistry. This deals with systems bigger than a single molecule – for instance, a DNA double helix is a supramolecular assembly.

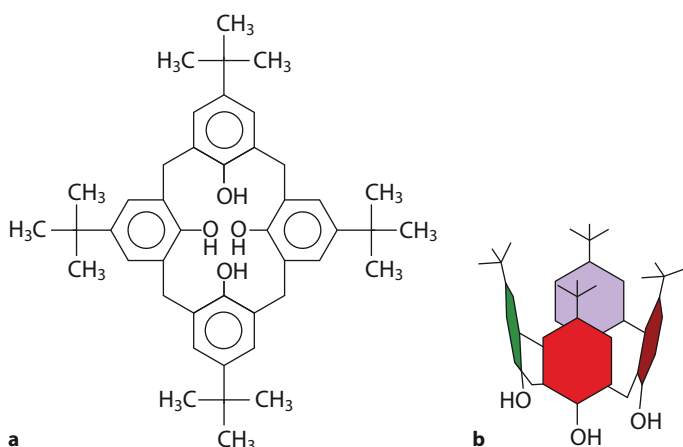


Figure B.46 **a** A calixarene molecule; **b** the cup-like structure of a calixarene – the rings have been shown in different colours to make the shape clearer.

Some uses of enzymes

Enzymes can be used to help the breakdown of oils spills, in the treatment of industrial waste and in biological detergents.

A major oil spill can have disastrous effects on the environment and many methods are used to try to control the extent of the spill, disperse it and remove the oil. One approach that has been proposed is the use of microbes to break down the oil in the spill.

Certain bacteria (and other microorganisms) possess enzymes that can break oil down – no one type of bacterium has the enzymes to break down *all* the components of crude oil and a mixture of microbes is required. These bacteria exist normally in the environment and will multiply rapidly when there is an oil spill. However, their ability to break oil down is limited by the availability of other nutrients, such as nitrogen and phosphorus, which they also need. It has been suggested that adding extra bacteria capable of breaking oil down, or a mixture of enzymes and extra nutrients for the bacteria already there, could be used to aid the breakdown of oil spills.

Immobilised enzymes (enzymes attached to a solid support) have also been used in the treatment of industrial waste water. They can break down specific chemicals, such as pesticides or cyanide, to prevent their release into the environment.

Enzymes are used in ‘biological’ washing powders. These enzymes – usually hydrolases – catalyse the breakdown of stains of fats (lipases), proteins (proteases) etc. Biological washing powders allow stain removal at lower temperatures than non-biological ones and so save energy. It is important that biological detergents are not used at high temperatures because the enzymes will be denatured and will not work.

Oil Spill Eater II (OSE II) is a commercial product containing enzymes that can convert crude oil into feedstock for naturally occurring bacteria and has been used in the treatment of oil spills.

Enzymes for industrial use are produced from microorganisms. These microorganisms are supplied with nutrients and kept under carefully controlled conditions. The microorganisms secrete enzymes, which are then separated from the reaction mixture.

Biological washing powders are designed to clean effectively at temperatures of 30 °C and below. They should not generally be used for washing at 60 °C and above.

Biomagnification

Biomagnification is the increase in concentration of a substance as it passes up a food chain.

For biomagnification to occur an organic substance should:

- not be broken down in the environment
- not be broken down in the body
- be lipid/fat-soluble so that it is not readily excreted but stored in fatty tissue instead.

One of the most famous cases of biomagnification is the pesticide DDT (dichlorodiphenyltrichloroethane – there are more systematic names!) (Figure B.47). This was used widely in agriculture for the control of insects and in the field of environmental health to kill insects that carry diseases such as malaria and typhus.

Some of the DDT used to spray areas of land found its way into rivers and lakes. It was taken up by microscopic plants, which were eaten by microscopic animals, which were eaten by fish, which were eaten by larger fish, which were eaten by birds (Figure B.48). At each level, the concentration of DDT in the organism increased. The effect of DDT on birds is a weakening of the shells of eggs so that they are unable to support the weight of the mother.

In a similar way, heavy metals, such as mercury, can become more concentrated as they move up a food chain and there have been worries about the level of mercury present in tuna consumed by humans, especially pregnant women. Sources of mercury include waste incineration, gold mining and coal combustion. Mercury finds its way into water supplies where it is absorbed by microorganisms, which convert it to methyl mercury (CH_3Hg^+). Its concentration increases as it is passed up

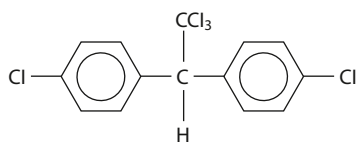


Figure B.47 DDT.



The use of DDT is a controversial issue and it was banned in most countries around the world. However, its use continued in some countries where it was used to control the mosquitoes that carry malaria. Malaria is one of the biggest killer diseases in the world and it is estimated that between half a million and a million people die annually from the disease. Recently (2013) African countries have decided to adopt DDT again to control mosquitoes which are resistant to other pesticides.

There are also health worries about biomagnification of many other organic substances – for example dioxins, which are produced as byproducts in the manufacture of some chlorinated organic compounds and from incineration of waste materials containing organochlorine compounds.

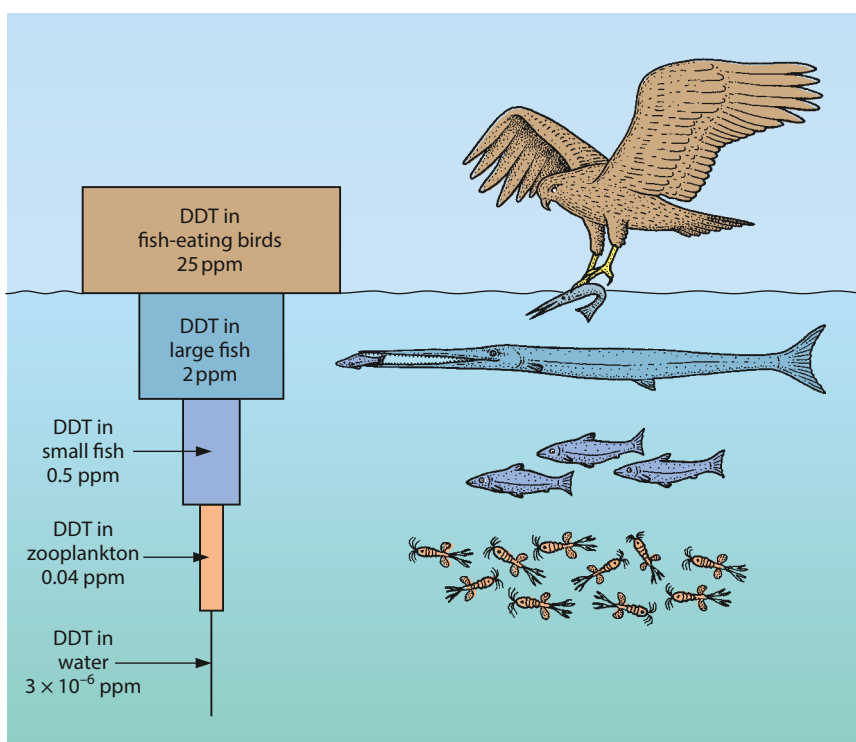


Figure B.48 DDT concentrations increase up the food chain.



the food chain until it reaches tuna (and other large fish such as swordfish and sharks). Adverse effects of consuming mercury include damage to the central nervous system.

Green chemistry

Green chemistry (also called ‘sustainable chemistry’) is an approach to chemical research and chemical industrial processes that seeks to minimise the production of hazardous substance and their release to the environment.

Paul Anastas, then of the United States Environmental Protection Agency, and John C. Warner developed the twelve principles of green chemistry.

These are:

- 1 *Prevention* – it is better to prevent waste than to treat or clean up waste after it has been created.
- 2 *Atom economy* – synthetic methods should be designed to maximise the incorporation of all materials used in the process into the final product.
- 3 *Less hazardous chemical syntheses* – wherever practicable, synthetic methods should be designed to use and generate substances that possess little or no toxicity to human health and the environment.
- 4 *Designing safer chemicals* – chemical products should be designed to affect their desired function, while minimising their toxicity.
- 5 *Safer solvents and auxiliaries* – the use of auxiliary substances (solvents, separation agents etc.) should be made unnecessary wherever possible and innocuous when used.
- 6 *Design for energy efficiency* – the energy requirements of chemical processes should be recognised for their environmental and economic impacts and should be minimised. If possible, synthetic methods should be conducted at ambient temperature and pressure.
- 7 *Use of renewable feedstocks* – a raw material or feedstock should be renewable, rather than depleting, whenever technically and economically practicable.
- 8 *Reduce derivatives* – unnecessary derivatisation (use of blocking groups, protection/deprotection, temporary modification of physical/chemical processes) should be minimised or avoided if possible because such steps require additional reagents and can generate waste.
- 9 *Catalysis* – catalytic reagents (as selective as possible) are superior to stoichiometric reagents.
- 10 *Design for degradation* – chemical products should be designed so that at the end of their function they break down into innocuous degradation products and do not persist in the environment.
- 11 *Real-time analysis for pollution prevention* – analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances.
- 12 *Inherently safer chemistry for accident prevention* – substances, and the form of a substance used in a chemical process, should be chosen to minimise the potential for chemical accidents including releases, explosions and fires.

(Green Chemistry: Theory and Practice by Paul Anastas & John Warner (1998) Figure 4.1 from p. 30. By permission of Oxford University Press www.oup.com)

As can be seen, there are many things that have to be considered when, for instance, making a substance that is required for biochemical research. One important consideration is the concept of atom economy. Atom economy can be used as a measure of how efficient a particular reaction is in terms of converting as much of the starting materials as possible into useful products.

$$\text{atom economy} = \frac{\text{molar mass of desired product}}{\text{total molar mass of all reactants}} \times 100\%$$

We can use the preparation of 1-phenylethanone, which could be investigated as an enzyme inhibitor, as an example to illustrate how the equation is used.

Consider two different ways of making 1-phenylethanone ($\text{C}_6\text{H}_5\text{COCH}_3$), from 1-phenylethanol:



The atom efficiency for each process can be worked out as follows.

Method 1:

$$\begin{aligned} \text{Total molar mass of all reactants} &= (3 \times 122.18) + (2 \times 100.00) + (3 \times 98.09) \\ &= 860.81 \text{ g mol}^{-1} \end{aligned}$$

$$\text{Molar mass of desired product} = 3 \times 120.16 = 360.48 \text{ g mol}^{-1}$$

$$\text{Atom economy} = \left(\frac{360.48}{860.81} \right) \times 100 = 41.88\%$$

Method 2:

$$\text{Total molar mass of all reactants} = 122.18 + (0.5 \times 32.00) = 138.18 \text{ g mol}^{-1}$$

$$\text{Molar mass of desired product} = 120.16 \text{ g mol}^{-1}$$

$$\text{Atom economy} = \left(\frac{120.16}{138.18} \right) \times 100 = 86.96\%$$

It can be seen that method 2 has a much higher atom economy and is, therefore, much more efficient. However, many other things must be considered when assessing these reactions in terms of green chemistry principles – the temperature used in each reaction, the solvents used and how much of them, disposing of the solvents, the nature of the catalyst required for the second reaction etc.

Nature of science

An understanding of science is essential if the public and politicians are to make informed judgments about the advantages and disadvantages of using substances such as DDT. It is essential that scientists provide the evidence in a way that is as complete as possible, but also objective so that people can make their own decisions. In the case of the re-introduction of DDT into Africa, the scientists provide the evidence but the politicians ultimately make the decisions.

Atom economy is not the same as the yield of a reaction. Atom economy is a theoretical quantity based on a chemical equation and allows evaluation of how much waste (products) will be produced.

The yield of a reaction is an experimental quantity worked out from how much of the desired product is actually made in a chemical reaction.

In the calculation of atom economy above, it has been assumed that all reactions have 100% yield, which will not be the case in practice.

When evaluating how green/ environmentally friendly a particular process is, both atom economy and yield must be considered – as well as several other factors.

When DDT was first introduced scientists did not consider possible negative effects on the environment but nowadays we are much more aware of such issues. When substances are made the environmental impact of the synthesis and use of the substance are often major considerations. There are two major approaches to the environmental issues and these are cure and prevention. In the past scientists have worked to develop a solution to a problem (possibly of their own creation) after it has arisen but prevention is now becoming a more important factor.

? Test yourself

23 Calculate the atom economy for each of the following reactions:

- $\text{CaC}_2 + \text{H}_2\text{O} \rightarrow \text{C}_2\text{H}_2 + \text{CaO}$, where the desired product is ethyne.
- $\text{C}_2\text{H}_4 + \text{PdCl}_2 + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{CHO} + \text{Pd} + 2\text{HCl}$, where the desired product is ethanal.
- $4\text{HgS} + \text{CaO} \rightarrow 4\text{Hg} + 3\text{CaS} + \text{CaSO}_4$, where the desired product is mercury.

B7 Proteins and enzymes (HL)

Enzyme kinetics

As discussed in Section B2, enzymes are proteins that catalyse biochemical reactions. A Michaelis–Menten curve (Figure B.49) shows how the rate of an enzyme-catalysed reaction varies with substrate concentration.

At low substrate concentrations, the rate of an enzyme-catalysed reaction is proportional to the substrate concentration – essentially it is first order with respect to substrate concentration. As the substrate concentration increases, the rate of reaction increases to a lesser extent and is no longer proportional to substrate concentration. At high substrate concentrations, the rate of reaction remains constant and does not increase further with an increase in substrate concentration – the rate of reaction is zero order with respect to the substrate concentration.

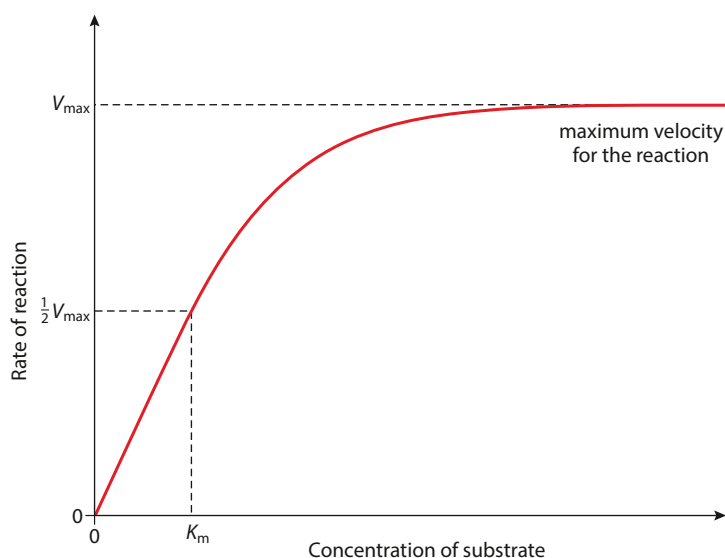


Figure B.49 A Michaelis–Menten curve – the concentration of enzyme, temperature, pH etc. are kept constant and the concentration of the substrate is varied.

Learning objectives

- Determine V_{max} and the Michaelis constant (K_m) by graphical means and explain the importance of K_m
- Compare the modes of action of competitive and non-competitive inhibitors, including how they affect V_{max} and K_m
- Understand how the product of an enzyme-catalysed reaction can act as an inhibitor
- Understand that amino acids and proteins can act as buffers in solution
- Calculate the pH of buffer solutions
- Understand how UV–Vis spectroscopy can be used for protein assays

Exam tip

V_{\max} has units of rate – e.g. $\text{mol dm}^{-3} \text{s}^{-1}$

K_m has units of concentration – e.g. mol dm^{-3}

The rate of reaction for an enzyme saturated with substrate is known as the **maximum velocity** (V_{\max}). V_{\max} varies from one enzyme to another and is dependent on reaction conditions such as temperature and pH.

The Michaelis constant (K_m) is the concentration of substrate when the rate is equal to one half of V_{\max} .

K_m is a useful concept because it gives an indication of the affinity of an enzyme for a substrate – how well the enzyme binds to the substrate (so it is a measure of the stability of the enzyme–substrate complex). If an enzyme has a low K_m , this indicates that it has a high affinity for a substrate because only a small concentration of substrate is needed for the reaction to proceed at half its maximum velocity. Large K_m values indicate that an enzyme has less affinity for the substrate because a large concentration of substrate is needed to reach half V_{\max} .

Enzyme inhibitors

If a chemical binds to an enzyme and prevents it from carrying out its catalytic activity, the enzyme is said to be **inhibited**, and the chemical is called an **enzyme inhibitor**. There are two main types of enzyme inhibitors, depending on where they interact with the enzyme – competitive inhibitors and non-competitive inhibitors. Enzyme inhibitors are widely used as medicinal drugs, the most common being competitive inhibitors.

Competitive enzyme inhibitors

As their name suggests, competitive inhibitors compete with the natural substrate for binding to the enzyme – they bind to the active site of the enzyme. Competitive inhibitors normally have a structure similar to the natural substrate – this allows them to form interactions with the active site. Once an inhibitor enters the active site, it binds to form an enzyme–inhibitor complex (rather than an enzyme–substrate complex). The inhibitor is not acted on by the enzyme and so does not form products – instead, it blocks the entry of substrate and stops the enzyme from acting on the substrate and carrying out catalysis (Figure B.50).

Competitive inhibition is usually reversible, so the enzyme–inhibitor complex breaks down to release the inhibitor from the active site. Because the inhibitor and substrate are competing for the active site, increasing the substrate concentration will make it more likely that the substrate

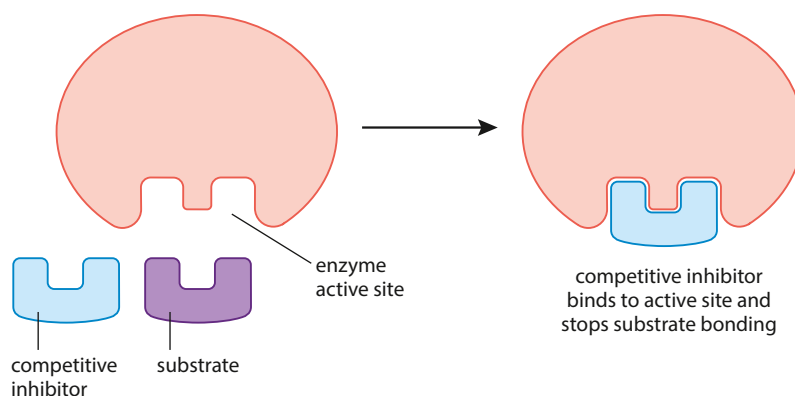


Figure B.50 The mechanism of a competitive inhibitor.

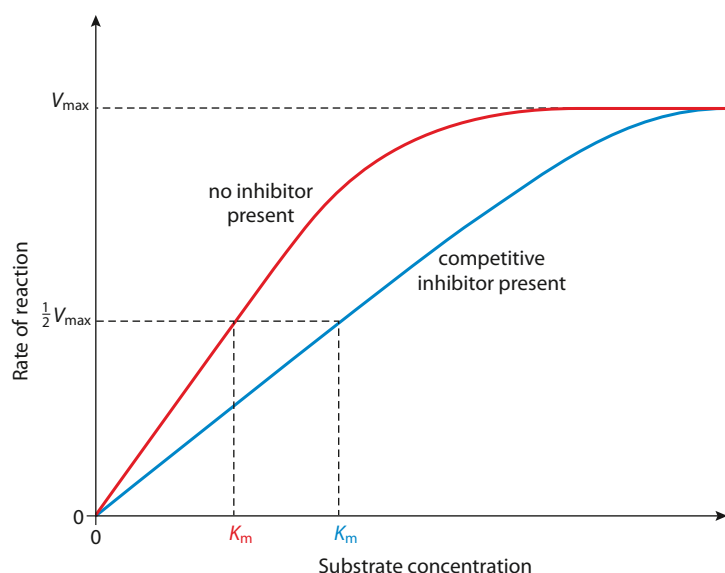


Figure B.51 The effect of a competitive inhibitor on V_{\max} and K_m .

will enter the active site, and inhibition will be reduced. The maximum velocity of the reaction will remain the same, but it will take a higher concentration of substrate to reach $\frac{1}{2}V_{\max}$, so the K_m will be higher (Figure B.51).

Non-competitive inhibitors

Non-competitive inhibitors do not compete with the natural substrate, because they do not bind to the active site but another region of the enzyme – an **allosteric** site. This binding causes a change in the shape of the active site, which prevents the substrate from binding (Figure B.52).

As the inhibitor does not compete with the substrate for the same site, increasing the substrate concentration does not reduce inhibition. This means that V_{\max} is reduced in the presence of this type of inhibitor, but that K_m is the same (Figure B.53). Adding a non-competitive inhibitor is essentially lowering the concentration of the enzyme available to catalyse reactions and so V_{\max} will decrease but it does not affect the ability of the remaining enzymes to bind substrate and so K_m is the same.

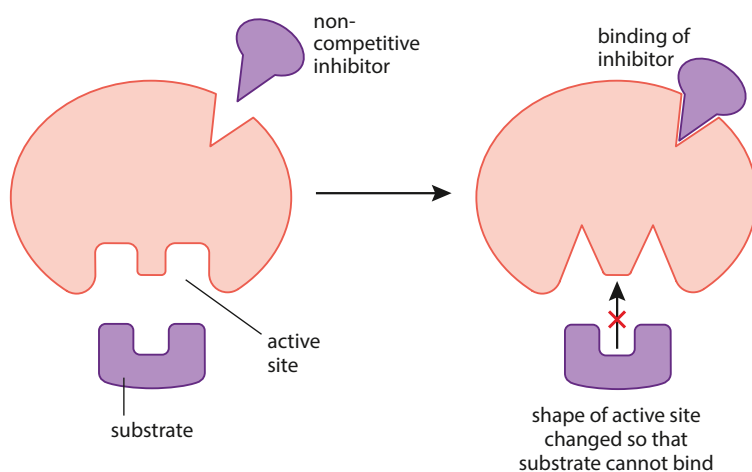


Figure B.52 The mechanism of a non-competitive inhibitor.

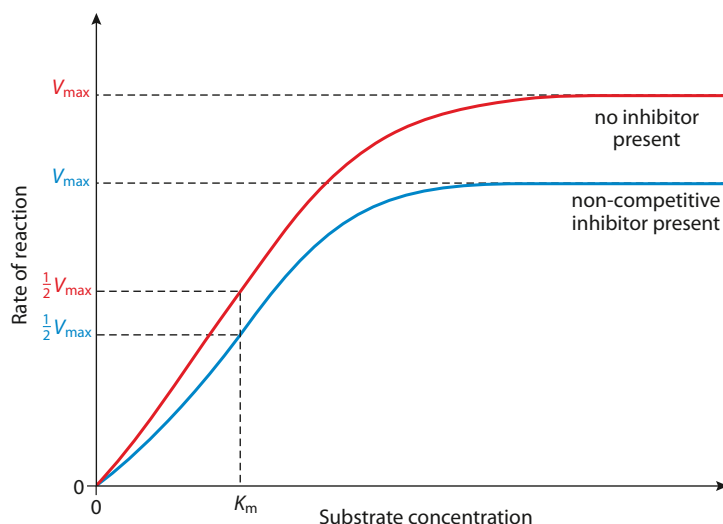


Figure B.53 The effect of a non-competitive inhibitor on V_{\max} and K_m .

Feedback control of enzymes

In feedback control of enzyme activity, a product of a particular metabolic pathway inhibits an enzyme earlier in the pathway. In this way the amount of product can be controlled. Consider the sequence of reactions shown in Figure B.54 for the conversion of the amino acid threonine into isoleucine where each step is controlled by an enzyme.

Isoleucine inhibits enzyme 1 in this sequence. When the concentration of isoleucine gets too high, the enzyme is inhibited more, which slows down the rate of production. When the concentration of isoleucine falls, the amount of inhibition decreases and the rate of production increases.

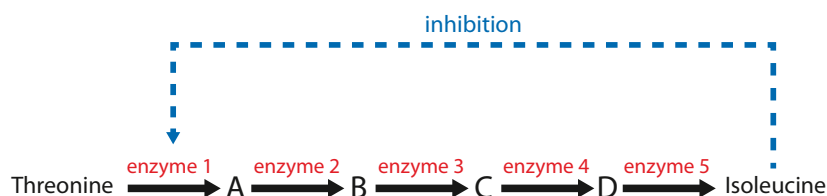
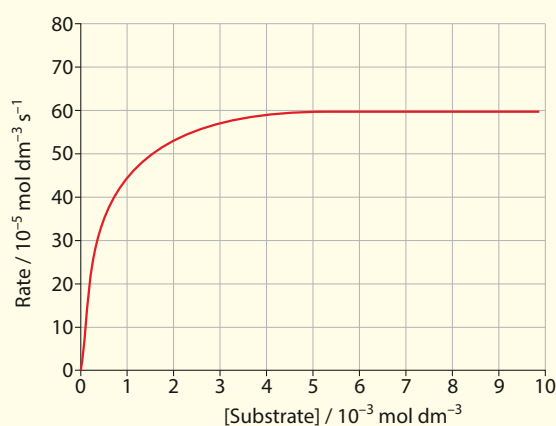


Figure B.54 Feedback inhibition.

? Test yourself

- 24 Work out V_{\max} and K_m from the graph.
- 25 Enzyme **X** has a K_m of $2.0 \times 10^{-7} \text{ mol dm}^{-3}$ for a particular substrate, whereas enzyme **Y** has a K_m of $2.0 \times 10^{-6} \text{ mol dm}^{-3}$ for the same substrate. Which enzyme has a greater affinity for the substrate?

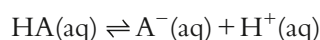


Buffer solutions

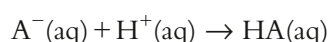
We have already met the idea of a buffer solution in Topic 8.

A buffer solution is one that resists changes in pH when small amounts of acid or alkali are added.

A buffer solution consists of two components – an acid and a base. The base reacts with any acid added and the acid reacts with any base added. Consider a general buffer containing acid HA and base A⁻. The equilibrium that exists in this solution is:

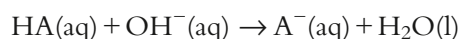


If some hydrochloric acid is added to this solution, the extra H⁺ reacts with the A⁻ (base) in the solution:



The H⁺ added is ‘mopped up’ by reaction with the base and therefore the pH changes very little.

If some sodium hydroxide is added to the solution, the extra OH⁻ reacts with the HA (acid) in the solution:



The OH⁻ added is ‘mopped up’ by reaction with the acid and, once again, the pH changes very little.

Amino acids and proteins can act as buffers

Amino acids play an important role in buffering the aqueous environment in cells. Significant changes in cell pH can have a disastrous effect on the biochemical reactions that take place there – they prevent enzymes, which usually only work within narrow pH ranges, from carrying out their catalytic activity. Other proteins may also change shape in low pH or high pH and lose their ability to function.

We mentioned earlier that amino acids are amphoteric – they contain both acidic and basic groups and can therefore act as either an acid or a base.

The zwitterionic form of an amino acid is shown in Figure B.55.

At its isoelectric point, an amino acid is in its zwitterionic form, but if some acid is added some of the zwitterion will be protonated and there will be an equilibrium mixture of the two species (Figure B.56). This system can then act as a buffer because there is an acid and a base present (X is the base because it can accept a proton and Y is the acid because it can donate a proton).

When some acid is added to this equilibrium mixture (Figure B.57) it

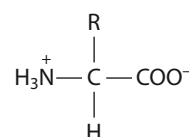


Figure B.55 A zwitterion.

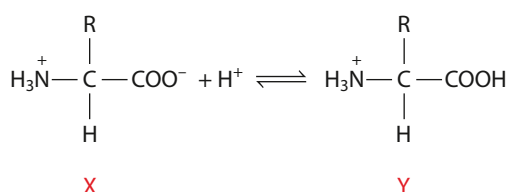


Figure B.56 Creating a buffer solution.

the number of moles of acid (H⁺) added is less than that of the amino acid.

is mopped up by reaction with **X**.

The added H^+ is therefore mostly removed from the solution and the

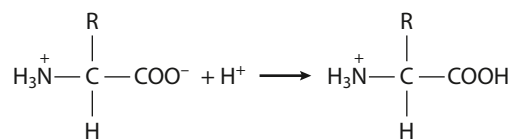


Figure B.57 Mopping up added acid.

pH does not change very much.

When some alkali is added to the equilibrium mixture (Figure **B.58**) it reacts with **Y**.

Therefore the added OH^- is mopped up.

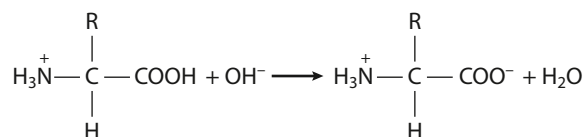


Figure B.58 Mopping up added alkali.

When some alkali is added to an amino acid that is at its isoelectric point, a different equilibrium mixture exists (Figure **B.59**), which can also act as a buffer:

Again there is an acid (**P**) and a base (**Q**) present and this mixture can

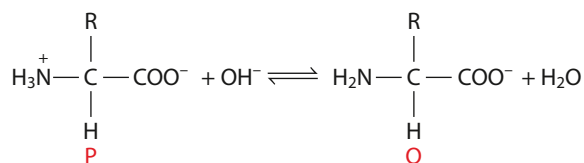


Figure B.59 A different equilibrium mixture – still a buffer.

also act as a buffer.

It is important to realise that an amino acid does not act as a buffer around its isoelectric point because there is only one species present. The titration curve for glycine (isoelectric point = 6) is shown in Figure **B.60**, with the buffering regions marked.

The situation is further complicated by the presence of side groups, some of which can also act as acids and bases.

Proteins can also act as buffers because of the side chains. Consider a protein that is rich in lysine (Figure **B.61**).

The amount of alkali added must be less than the amount of amino acid present.

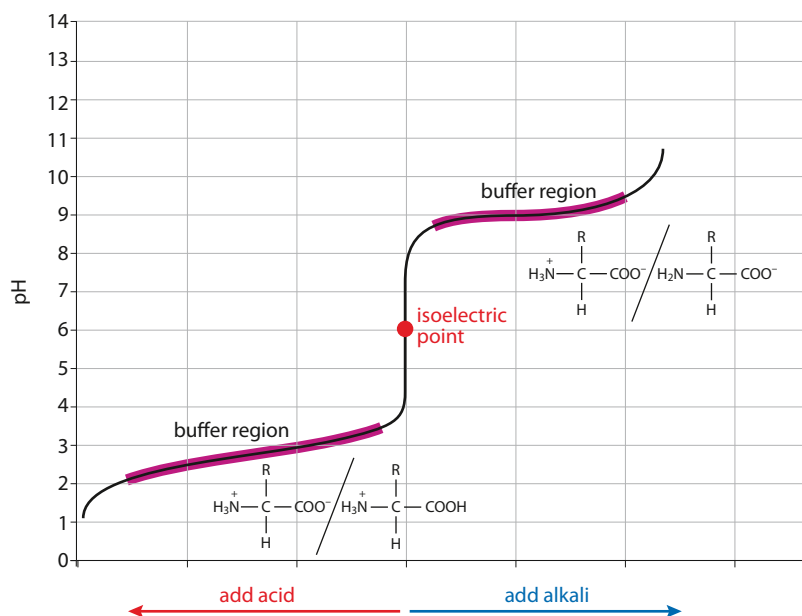


Figure B.60 The buffering action of an amino acid.

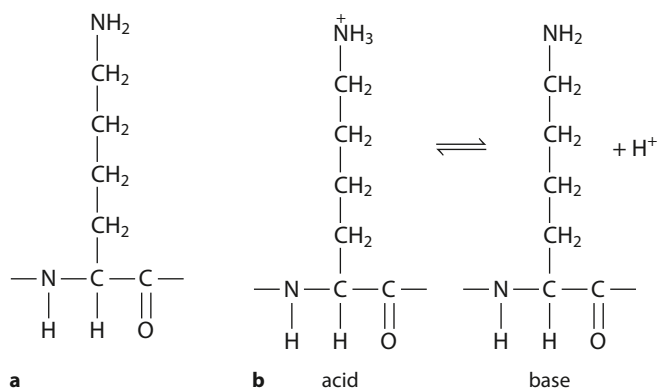


Figure B.61 **a** Part of a protein chain showing the side chain in a lysine residue; **b** an equilibrium mixture containing significant amounts of acid and base can act as a buffer.

At pH 10, some of the NH_2 side groups will be protonated and others will not (Figure B.61b). So there will be a mixture of acid and base present – and therefore a buffer.

How to calculate the pH of a buffer solution

Buffers are used extensively in biochemical research to control the pH during biochemical reactions.

Worked example

B.5 Calculate the pH of a solution containing $0.200 \text{ mol dm}^{-3}$ ethanoic acid ($K_a = 1.74 \times 10^{-5} \text{ mol dm}^{-3}$) and $0.250 \text{ mol dm}^{-3}$ sodium ethanoate.

The equilibrium that exists in this solution is:



K_a for this equilibrium is given by:

$$K_a = \frac{[\text{CH}_3\text{COO}^-(\text{aq})][\text{H}^+(\text{aq})]}{[\text{CH}_3\text{COOH}(\text{aq})]}$$

We will make the approximation that the equilibrium concentrations of ethanoic acid and ethanoate ions are the same as their initial concentrations. This is a reasonable assumption because the dissociation of ethanoic acid in pure water is very low, and therefore the dissociation is going to be even lower if some CH_3COO^- is already present (Le Chatelier's principle – the position of the dissociation equilibrium shifts to the left if CH_3COO^- is added).

We can now substitute values into the K_a expression:

$$1.74 \times 10^{-5} = \frac{0.250 \times [\text{H}^+(\text{aq})]}{0.200}$$

Rearranging and calculating:

$$[\text{H}^+(\text{aq})] = 1.39 \times 10^{-5} \text{ mol dm}^{-3}$$

$$\text{pH} = -\log_{10}[\text{H}^+(\text{aq})]$$

$$= -\log_{10}(1.39 \times 10^{-5})$$

$$= 4.86$$

The pH of this buffer solution is 4.86.

Sodium ethanoate is a salt and breaks apart completely into its ions in aqueous solution. The concentration of ethanoate ions is therefore the same as that of sodium ethanoate.

The Henderson–Hasselbalch equation

For a buffer solution made up of a mixture of HA (acid) and A^- (base), the pH of the buffer can be worked out using the Henderson–Hasselbalch equation:

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{A}^-]}{[\text{HA}]}} \right)$$

Another way of writing this is:

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}} \right)$$

A common buffer used in the study of biochemical reactions is a phosphate buffer containing dihydrogen phosphate ions (H_2PO_4^-) and hydrogen phosphate (HPO_4^{2-}) ions. The equilibrium that exists in this solution is:



H_2PO_4^- has an extra proton and acts as an acid, whereas HPO_4^{2-} with one fewer proton acts as a base.

Exam tip

The species with more hydrogen will be the acid (HA); the species with fewer hydrogen/more negative charge/less positive charge will be the base (A^-).

Worked examples

B.6 Calculate the pH of a buffer solution containing $0.0550 \text{ mol dm}^{-3} \text{ H}_2\text{PO}_4^-$ ($\text{p}K_a = 7.21$) and $0.0450 \text{ mol dm}^{-3} \text{ HPO}_4^{2-}$.

$$[\text{base}] = 0.0450 \text{ mol dm}^{-3}; [\text{acid}] = 0.0550 \text{ mol dm}^{-3}$$

$$\begin{aligned}\text{pH} &= \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right) \\ &= 7.21 + \log_{10} \left(\frac{0.0450}{0.0550} \right) \\ &= 7.21 + \log_{10} 0.818 \\ &= 7.21 - 0.0872 \\ &= 7.12\end{aligned}$$

B.7 Calculate the pH of a solution containing $0.400 \text{ mol dm}^{-3}$ ammonia ($\text{p}K_b = 4.75$) and $0.200 \text{ mol dm}^{-3}$ ammonium chloride.

The Henderson–Hasselbalch equation can be used here. Ammonia (NH_3) is the base and the ammonium ion (NH_4^+) produced when ammonium chloride dissolves in water is the acid (has an extra H^+). The chloride ion is not important in this system – it is just the counter ion to balance out the charge on the ammonium ion.

In the Henderson–Hasselbalch equation we need the $\text{p}K_a$ of the acid and this can be worked out from the $\text{p}K_b$ of the base.

For a conjugate acid–base pair $\text{p}K_a + \text{p}K_b = \text{p}K_w$; at 25°C , $\text{p}K_a + \text{p}K_b = 14$

Therefore, $\text{p}K_a + 4.75 = 14$ and $\text{p}K_a$ for NH_4^+ is 9.25

Now $[\text{base}] = 0.400 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.200 \text{ mol dm}^{-3}$

$$\begin{aligned}\text{pH} &= \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right) \\ \text{pH} &= 9.25 + \log_{10} \left(\frac{0.400}{0.200} \right) \\ &= 9.25 + \log_{10} 2.00 \\ &= 9.25 + 0.301 \\ &= 9.55\end{aligned}$$

Calculating the pH of a buffer solution when volumes are given

Worked examples

B.8 A buffer solution is formed when 30.0 cm^3 of 0.100 mol dm^{-3} potassium dihydrogen phosphate (KH_2PO_4) is added to 40.0 cm^3 of 0.110 mol dm^{-3} disodium hydrogen phosphate (Na_2HPO_4). $\text{p}K_a$ for H_2PO_4^- is 7.21. Calculate the pH of the mixture.

The first step is to work out the concentrations of the acid and base in the buffer solution.

The total volume of the solution is 70.0 cm^3 . Because the same number of moles of potassium dihydrogen phosphate are now present in 70.0 cm^3 instead of 30.0 cm^3 , the concentration of the potassium dihydrogen phosphate has decreased by a factor of $\frac{30.0}{70.0}$

The concentration of potassium dihydrogen phosphate in this solution will be:

$$\left(\frac{30.0}{70.0}\right) \times 0.100 = 0.0429\text{ mol dm}^{-3}$$

The concentration of disodium hydrogen phosphate in this solution will be:

$$\left(\frac{40.0}{70.0}\right) \times 0.110 = 0.0629\text{ mol dm}^{-3}$$

$$\text{So } [\text{base}] = 0.0629\text{ mol dm}^{-3}; [\text{acid}] = 0.0429\text{ mol dm}^{-3}$$

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]} \right)$$

$$\text{pH} = 7.21 + \log_{10} \left(\frac{0.0629}{0.0429} \right)$$

$$= 7.21 + \log_{10} 1.47$$

$$= 7.21 + 0.166$$

$$= 7.38$$

More figures were carried through on the calculator to give this answer.

Exam tip

When working out the pH of a buffer solution, you can check whether or not your answer is reasonable. If the solution contains a higher concentration of acid than base, the pH of the solution will be lower than the $\text{p}K_a$ of the acid; if there is a higher concentration of base than acid, the pH will be higher than $\text{p}K_a$.

The concentration of each species in the buffer solution can also be worked out using a moles calculation.

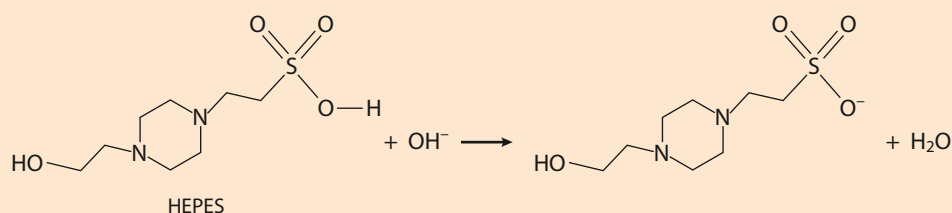
The number of moles of potassium dihydrogen phosphate in 30.0 cm^3 :

$$\left(\frac{30.0}{1000}\right) \times 0.100 = 0.00300\text{ mol}$$

So the concentration of potassium dihydrogen phosphate in the buffer solution is:

$$\left(\frac{0.00300}{70}\right) \times 1000 = 0.0429\text{ mol dm}^{-3}$$

B.9 HEPES is used in some biological buffers. A buffer solution can be made by dissolving sodium hydroxide in a HEPES solution.



Calculate the pH of the buffer solution formed when 20.0 g of sodium hydroxide is added to 1.00 dm³ of a 1.00 mol dm⁻³ solution of HEPES ($pK_a = 7.5$). Assume that there is no change in volume when the sodium hydroxide is added.

HEPES has an 'extra' proton and is therefore an acid. Reaction with sodium hydroxide converts some of it into a base.

M_r for sodium hydroxide is 40.00

So the number of moles of sodium hydroxide = $\frac{20.0}{40.00} = 0.500$ mol.

From the equation, there is a 1:1 reaction with sodium hydroxide and therefore 0.500 mol HEPES reacts with 0.500 mol NaOH to form 0.500 mol of the anion.

In 1.00 dm³ of a 1.00 mol dm⁻³ solution of HEPES there is 1.00 mol of HEPES. So if 0.500 mol react there will be 0.500 mol remaining. Therefore the concentration of HEPES and the anion in the buffer solution are both equal at 0.500 mol dm⁻³.

$[\text{base}] = 0.500 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.500 \text{ mol dm}^{-3}$

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}} \right)$$

$$\text{pH} = 7.5 + \log_{10} \left(\frac{0.500}{0.500} \right)$$

$$= 7.5 + \log_{10} 1$$

$$= 7.5 + 0$$

$$= 7.5$$

Determining the composition of a buffer solution given its pH

Worked example

B.10 A student wants to make up a buffer solution at pH 7.7 using $0.100 \text{ mol dm}^{-3}$ solutions of HEPES ($\text{p}K_{\text{a}}=7.5$) and its sodium salt. Calculate how much of each solution must be used to make 500 cm^3 of a buffer of pH 7.7.

We need to calculate the ratio of the acid and base in the buffer solution – this can be worked out using the Henderson–Hasselbalch equation.

$$\text{pH} = \text{p}K_{\text{a}} + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$7.7 = 7.5 + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$\log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right) = 0.2$$

$$\frac{\text{base}}{\text{acid}} = 10^{0.2}$$

$$= 1.58$$

To get rid of 'log₁₀' use the inverse function – use the shift/2nd + log key combination on your calculator

Therefore the ratio [base]:[acid] is 1.58:1

Because the concentrations of the solutions are the same, the amount of each solution required to make 500 cm^3 of buffer can be worked out as:

$$\text{volume of base} = \left(\frac{1.58}{2.58} \right) \times 500 = 306 \text{ cm}^3$$

$$\text{volume of acid} = \left(\frac{1.00}{2.58} \right) \times 500 = 194 \text{ cm}^3$$

2.58 is just $1.58 + 1$ from the ratio.

This could also be worked out using $500 - 306$.

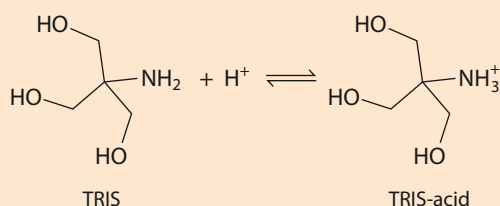
Therefore the volume of the HEPES solution required is 194 cm^3 and that of the solution of its sodium salt is 306 cm^3 .

If all figures are carried through on the calculator the answers 193 cm^3 and 307 cm^3 are obtained.

Calculating the change in pH of a buffer solution when acid or alkali is added

Worked example

B.11 TRIS is used as a buffer in biochemistry. A buffer solution is prepared by adding hydrochloric acid to TRIS to form a mixture of TRIS and its protonated form (TRIS-acid). The equilibrium that exists in the buffer solution is:



a Calculate the pH of a buffer solution containing $0.750 \text{ mol dm}^{-3}$ TRIS-acid ($\text{p}K_{\text{a}} = 8.30$) and $0.750 \text{ mol dm}^{-3}$ TRIS.

b What is the pH of the solution formed when 10.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ hydrochloric acid is added to 50.0 cm^3 of the buffer solution in part **a**?

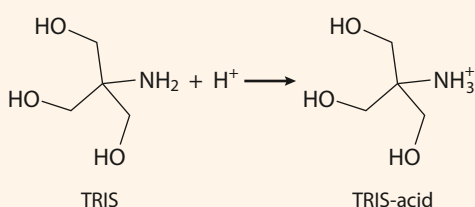
a $[\text{base}] = 0.750 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.750 \text{ mol dm}^{-3}$

$$\begin{aligned}\text{pH} &= \text{p}K_{\text{a}} + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right) \\ &= 8.30 + \log_{10} \left(\frac{0.750}{0.750} \right) \\ &= 8.30\end{aligned}$$

Exam tip

When $[\text{base}] = [\text{acid}]$, the pH of the buffer is equal to the $\text{p}K_{\text{a}}$ of the acid.

b When some acid is added to the buffer solution, the following reaction occurs:



This means that the concentration of TRIS decreases and that the concentration of TRIS-acid increases. To work out by how much they change we need to work out the initial number of moles of TRIS and TRIS-acid and how many moles of acid were added.

The number of moles of TRIS in 50.0 cm^3 of $0.750 \text{ mol dm}^{-3}$ solution is given by:

$$\begin{aligned}\text{no. moles} &= \text{concentration} \times \text{volume in dm}^3 \\ &= 0.750 \times \left(\frac{50.0}{1000} \right) \\ &= 0.0375 \text{ mol}\end{aligned}$$

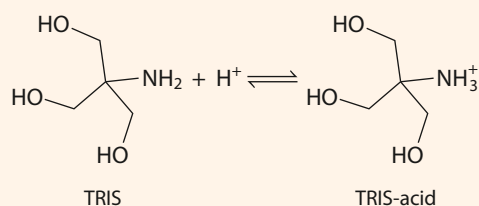
This is the same as the number of moles of TRIS-acid.

The number of moles of hydrochloric acid in 10.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ solution is given by:

$$\begin{aligned}\text{no. moles} &= 0.100 \times \left(\frac{10.0}{1000} \right) \\ &= 1.00 \times 10^{-3} \text{ mol}\end{aligned}$$

We will assume that the H^+ from the hydrochloric acid reacts with the equivalent amount of TRIS and that there is no further change in the number of moles as equilibrium is established.

So the number of moles of TRIS decreases by $1.00 \times 10^{-3} \text{ mol}$ and the number of moles of TRIS-acid increases by $1.00 \times 10^{-3} \text{ mol}$.



Initial amounts / mol:	0.0375	0.0375 mol
After HCl added / mol:	$0.0375 - 1.00 \times 10^{-3}$	$0.0375 + 1.00 \times 10^{-3}$
	0.0365 mol	0.0385 mol

The concentration of each species can be worked out by dividing the number of moles by the total volume in dm^3 , which is $50.0 + 10.0 = 60.0 \text{ cm}^3$ or 0.0600 dm^3

Concentration/ mol dm^{-3} :	$\frac{0.0365}{0.0600}$	$\frac{0.0385}{0.0600}$
	0.608	0.642

So $[\text{base}] = 0.608 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.642 \text{ mol dm}^{-3}$

$$\begin{aligned}
 \text{pH} &= \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]} \right) \\
 &= 8.30 + \log_{10} \left(\frac{0.608}{0.642} \right) \\
 &= 8.28
 \end{aligned}$$

So, on addition of 10.0 cm^3 of the hydrochloric acid, the pH of the buffer solution falls by 0.02 to 8.28.

? Test yourself

- 26** Calculate the pH values of the following buffer solutions.
- A solution containing $0.0200 \text{ mol dm}^{-3}$ butanoic acid ($\text{p}K_a = 4.82$) and $0.0200 \text{ mol dm}^{-3}$ sodium butanoate.
 - A solution containing $0.0500 \text{ mol dm}^{-3}$ propanoic acid ($\text{p}K_a = 4.87$) and $0.0200 \text{ mol dm}^{-3}$ sodium propanoate.
 - A solution containing $0.300 \text{ mol dm}^{-3}$ ethanoic acid ($\text{p}K_a = 4.76$) and $0.500 \text{ mol dm}^{-3}$ sodium ethanoate.
 - A solution made up by mixing together 25.0 cm^3 of $0.200 \text{ mol dm}^{-3}$ ethanoic acid ($\text{p}K_a = 4.76$) and 50 cm^3 of $0.100 \text{ mol dm}^{-3}$ sodium ethanoate.
 - A solution obtained when 10.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ sodium hydroxide is added to 20.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ ethanoic acid ($\text{p}K_a = 4.76$).
 - A solution obtained when 20.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ ammonia solution ($\text{p}K_b = 4.75$) is added to 40.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ ammonium chloride solution.
- 27**
- A buffer solution contains ethanoic acid ($\text{p}K_a = 4.76$), at a concentration of 1.00 mol dm^{-3} , and sodium ethanoate. If the pH of the buffer solution is 4.20, what is the concentration of the sodium ethanoate?
 - 20.0 cm^3 of $0.0100 \text{ mol dm}^{-3}$ hydrochloric acid is added to 50 cm^3 of the buffer solution in part **a**. Calculate the new pH of the buffer solution.

Protein assays

A protein assay is a method of determining the concentration of a protein in solution. Protein assays usually involve ultraviolet–visible (UV–Vis) spectroscopy. There are two basic approaches – either the absorption of light by the protein in the UV region of the spectrum is measured, or a coloured dye is added which binds to the protein and the absorption in the visible region of the spectrum of the protein–dye complex is measured. In each case the concentration of the protein is determined by reference to a calibration curve which is constructed using known concentrations of protein.

UV spectroscopy

Proteins absorb electromagnetic radiation in the UV region due to the presence of aromatic rings in the side chains. The wavelength at which maximum absorbance occurs is 280 nm. The following steps must be followed to construct a calibration curve and determine the concentration of a protein solution:

- The UV–Vis spectrophotometer is zeroed with a cuvette (a cuboid container used in spectrophotometers/colorimeters) containing just the solvent.
- A series of protein solutions of known concentration are made up – for example, 1.00 mg dm^{-3} , 2.00 mg dm^{-3} , 3.00 mg dm^{-3} etc.
- The absorbance of each of these protein solutions is measured at 280 nm.
- A calibration curve of absorbance against concentration is plotted (Figure B.62).
- The absorbance of the protein solution of unknown concentration is measured at 280 nm using a UV–Vis spectrophotometer.
- The concentration of the unknown solution is read off the calibration curve.

If the absorbance of the unknown protein solution was 0.29 the concentration can be read off the calibration curve in Figure B.62 as 3.6 ppm.

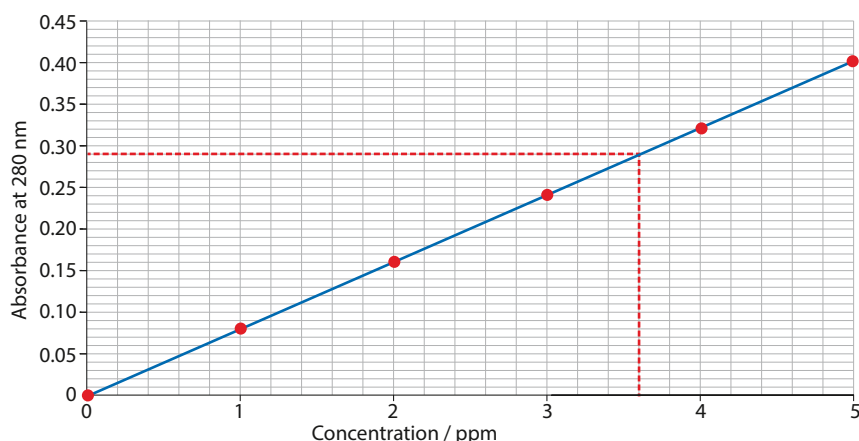


Figure B.62 A calibration curve for a protein assay using UV spectroscopy.

Visible spectroscopy

There are various methods for determining the concentration of protein using visible spectroscopy. The basic technique is the same as above in that a calibration curve must first be constructed using solutions of known

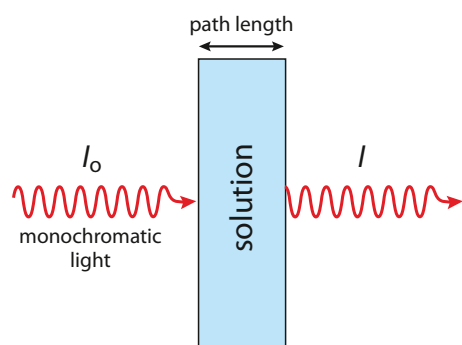


Figure B.63 The Beer–Lambert law relates the absorbance of light by a solution to the path length and concentration.

Absorbance has no units.

ϵ depends on the absorbing substance and on the wavelength of the light.

concentrations. One of the most common methods is the Bradford assay, which involves adding Bradford reagent to the protein sample. This reagent contains Coomassie Brilliant Blue dye which binds to the protein. The bound and unbound forms of the dye have different colours and the concentration of the protein can be determined by measuring the absorbance at 595 nm, which is in the visible region of the spectrum.

The Beer–Lambert law

This law relates the amount of light absorbed by a solution to its concentration and the path length (Figure B.63).

The Beer–Lambert law is:

$$\log_{10}\left(\frac{I_0}{I}\right) = A = \epsilon cl$$

where:

I_0 is the intensity of the light before it passes through the sample

I is the intensity of the light after it has passed through the sample

A is the absorbance $= \log_{10}\left(\frac{I_0}{I}\right)$

ϵ is the molar absorptivity (units: $\text{cm}^{-1} \text{mol}^{-1} \text{dm}^3$)

c is the concentration of the solution in mol dm^{-3}

l is the path length – the thickness of the sample (usually in cm).

The Beer–Lambert law is usually used in the form: $A = \epsilon cl$

It tells us that more radiation is absorbed by a more concentrated solution or if the radiation has to pass through a thicker sample. If the light encounters twice as many particles as it passes through the sample, twice as much will be absorbed.

The concentration of protein in a sample can be determined from absorbance data and application of the Beer–Lambert law. First of all the value of ϵ must be worked out – this can be done either by measuring the absorbance of a solution of known concentration or from the calibration curve. If the data in Figure B.62 were obtained using a constant path length of 1 cm, we have:

$$A = \epsilon c$$

and so ϵ is the gradient of the graph of absorbance against concentration. From Figure B.62 we can work out the value of the molar absorptivity:

$$\begin{aligned}\epsilon &= \frac{0.32}{4.0} \\ &= 0.080 \text{ cm}^{-1} \text{ ppm}^{-1}.\end{aligned}$$

The concentration of our unknown solution, with absorbance 0.29 and path length 1.0 cm, can now be worked out:

$$\begin{aligned}A &= \epsilon cl \\ c &= \frac{A}{(\epsilon \times l)} \\ &= \frac{0.29}{(0.080 \times 1)} \\ &= 3.6 \text{ ppm}\end{aligned}$$

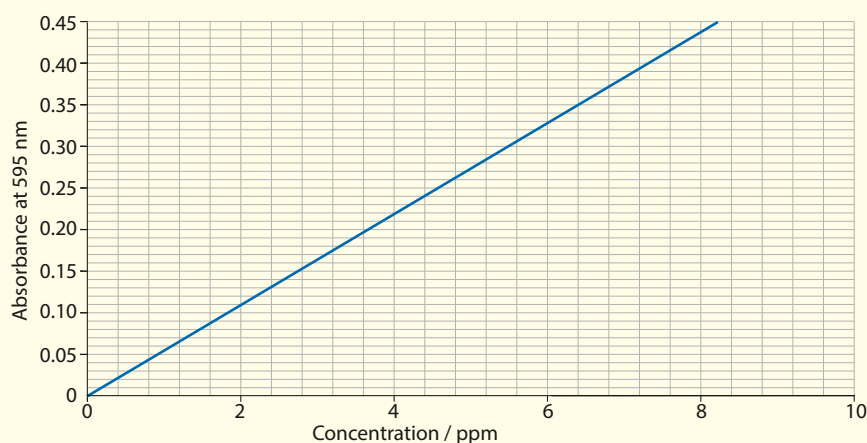
Nature of science

Protein analyses are used routinely in analytical chemistry, but it is important for scientists to have an understanding of the accuracy and precision of their data. They must be aware of possible systematic errors in their procedures which, although they might give reproducible results, are not accurate. Different protocols have been developed from the very simple methods described here to give reliable and accurate readings for protein concentrations.

Science is a highly collaborative field and collaboration between biochemists and organic chemists has led to significant advances in the use of enzymes in organic synthesis.

? Test yourself

- 28 A Bradford assay was carried out to measure the concentrations of some protein solutions. The calibration curve is shown below.



Determine the concentration in ppm of the protein solutions with the absorbances given below.

- a absorbance = 0.26
 - b absorbance = 0.15
 - c absorbance = 0.39
- 29 Calculate the concentration of each of the following protein solutions. All absorbances were measured at 280 nm in a cuvette of path length 1.0 cm.
- a the molar absorptivity at 280 nm for a particular protein solution is $500 \text{ cm}^{-1} \text{ mol}^{-1} \text{ dm}^3$ and the absorbance = 0.31
 - b the molar absorptivity at 280 nm for a particular protein solution is $63.5 \text{ cm}^{-1} \text{ mol}^{-1} \text{ dm}^3$ and the absorbance = 0.23
 - c the molar absorptivity at 280 nm for a particular protein solution is $1.02 \times 10^3 \text{ cm}^{-1} \text{ mol}^{-1} \text{ dm}^3$ and the absorbance = 0.18

Learning objectives

- Understand that nucleotides are the products of condensation reactions
- Understand that polynucleotides are formed in condensation reactions
- Understand the structure of DNA
- Understand how the structure of DNA differs from that of RNA
- Understand how to determine a nucleotide sequence in a complementary chain
- Understand that the sequence of bases in DNA determines the primary structure of proteins
- Understand what is meant by genetically modified foods
- Discuss some of the benefits and concerns of growing and eating GM foods

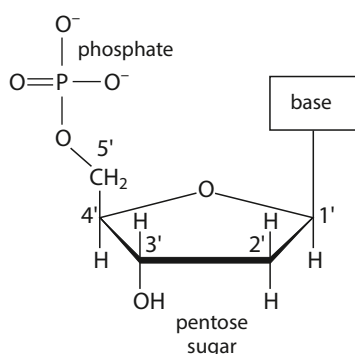


Figure B.64 The general structure of a nucleotide from DNA.

B8 Nucleic acids (HL)

As their names suggest, deoxyribonucleic acid (**DNA**) and ribonucleic acid (**RNA**) are nucleic acids. DNA carries the genetic code of an organism. When cells divide (make copies of themselves), the DNA of the cell must be copied (**DNA replication**) so that the new cells will have the same set of genetic information within them. When organisms reproduce, they copy their DNA and pass it on to the next generation. All organisms store their genetic information in DNA – except for some viruses that use RNA.

Nucleotides

Nucleic acids are polymers made up of monomers known as **nucleotides**. Nucleic acids are also called **polynucleotides**.

Nucleotides consist of a five-carbon (pentose) sugar, an organic nitrogen-containing heterocycle (nitrogen is part of the ring) called a base, and a phosphate group (Figure B.64).

The pentose sugar is different in DNA and RNA. In RNA the sugar is **ribose**, and in DNA the sugar is **2-deoxyribose**, so-called because it lacks an oxygen at the C2' position (Figure B.65).

Note that we use the numbering C1' to C5' when considering the sugar carbons in a nucleotide. This is because the atoms of the base in nucleotides are numbered 1 to 6 (in pyrimidines) or 1 to 9 (in purines), and so to avoid confusion the carbon atoms in the sugar are numbered C1' to C5'.

The bases of nucleic acids are derivatives of either **purine** or **pyrimidine** (Figure B.66) and are known as purines and pyrimidines.

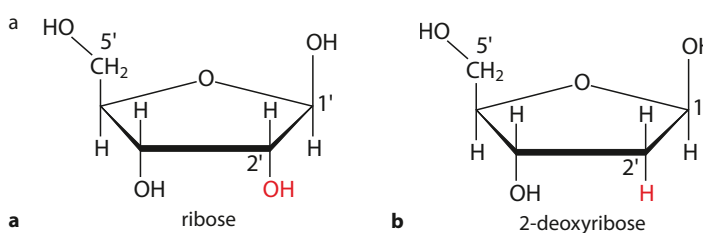


Figure B.65 The structures of **a** ribose and **b** 2-deoxyribose.

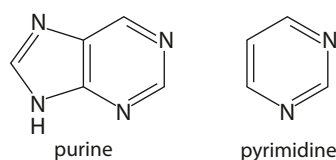


Figure B.66 The basic structure of purine and pyrimidine bases.

The purines that occur in nucleic acids are **adenine** (A) and **guanine** (G); the pyrimidines are **cytosine** (C), **thymine** (T, in DNA, not in RNA) and **uracil** (U, in RNA, not in DNA) (Figure B.67).

The base bonds to the sugar at the C1' position in a condensation reaction to form a **nucleoside** (Figure B.68).

The nucleoside can then undergo another condensation reaction with phosphoric acid/phosphate to form a nucleotide (Figure B.69). The phosphate group in a nucleotide is usually attached to the oxygen of the C5' hydroxyl group of the pentose sugar.

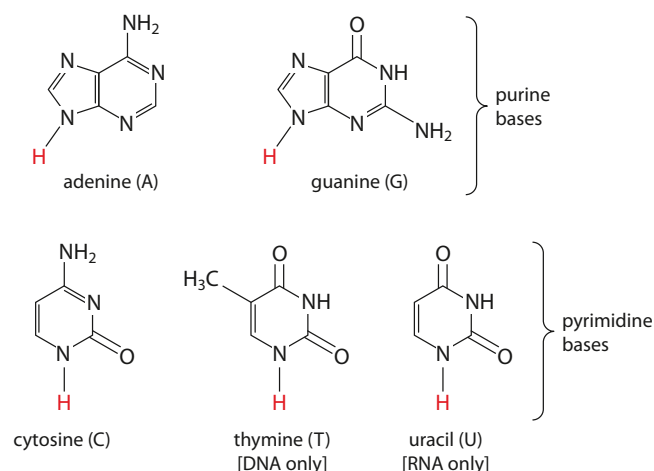


Figure B.67 The structures of purine and pyrimidine bases found in DNA and RNA; the hydrogen in red is the one that is lost when bases undergo a condensation reaction with a sugar molecule.

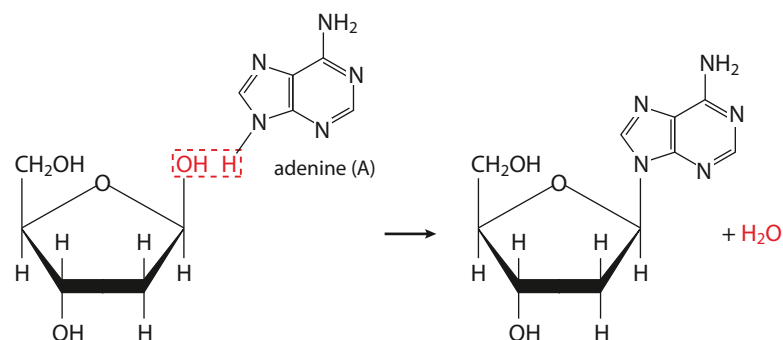


Figure B.68 Formation of a nucleoside in a condensation reaction.

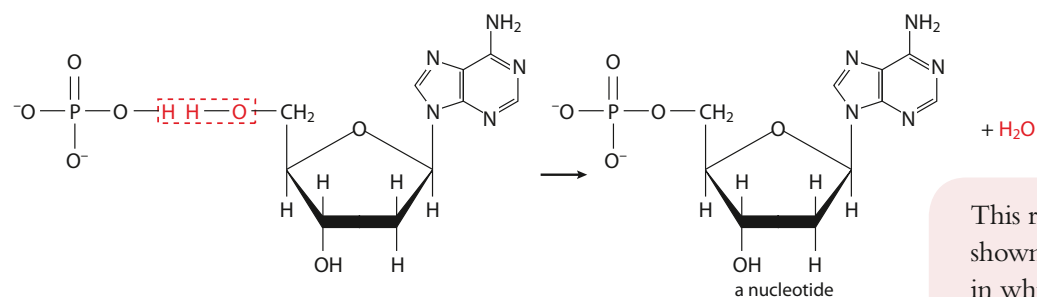
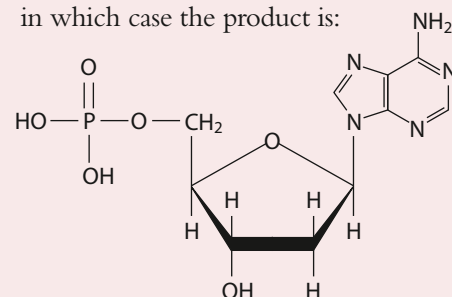


Figure B.69 Formation of a nucleotide from a nucleoside in a condensation reaction.

This reaction also could have been shown using phosphoric acid, H₃PO₄, in which case the product is:



Exam tip

You are not required to learn the structures of the nucleotide bases – but you should be able to recognise them.

Nucleotides join together to form polynucleotides. This is also a condensation reaction, between the phosphate attached to the C5' of one nucleotide and the hydroxyl group at C3' of another nucleotide. This results in the nucleotides being joined by a **phosphodiester link** (phosphate ester on both sides) between the sugars (Figure B.70). Nucleic acids may consist of thousands of nucleotides linked together.

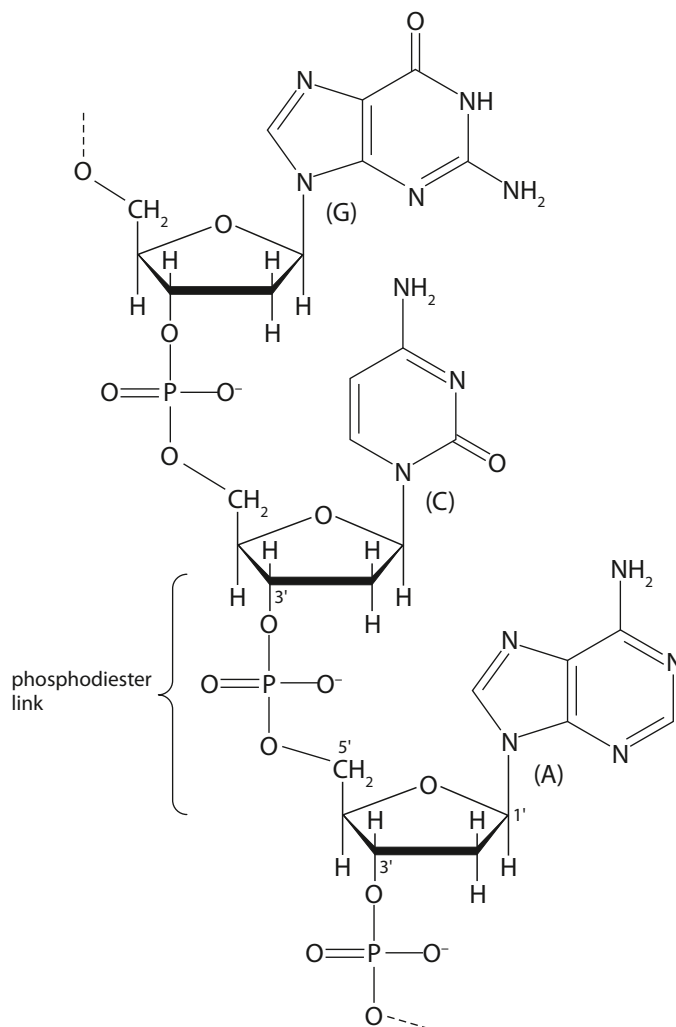


Figure B.70 Nucleotides joined by a phosphodiester link in a strand of DNA.

The presence of the phosphate groups in DNA means that it has a negative charge.

RNA is a single strand of nucleotides, whereas DNA is a double strand held together by hydrogen bonding. In both cases each strand of nucleotides has a repeating sugar–phosphate–sugar–phosphate backbone, and attached to each sugar is a base.

Differences between RNA and DNA

The differences between DNA and RNA are summarised as follows:

- DNA is a double-stranded nucleic acid, whereas RNA is single stranded.
- DNA contains thymine, whereas RNA contains uracil as a base.
- DNA contains 2-deoxyribose (no OH on C2') as the sugar, whereas RNA contains ribose.

Structure of DNA

Watson and Crick famously proposed the double-stranded structure of DNA in 1953 (Figure B.71). DNA is a double helix, consisting of two polynucleotide strands that spiral around an axis. Each complete turn of the helix is ten nucleotides in length.

The sugar–phosphate backbone in each strand winds around the outside of the helix, with the nucleotide bases attached to the sugars stacked in the interior of the helix. The hydrophilic sugar–phosphate backbone is exposed to the aqueous environment of the cell, while the relatively hydrophobic bases are shielded from this environment in the interior of the helix. These interactions stabilise the double-helix structure of DNA.

Each of the bases on one strand forms hydrogen bonds with a base on the opposite strand. The bases pair together in a specific way – the adenine bases on one strand form hydrogen bonds only with the thymine bases on the opposite strand (two hydrogen bonds), and the guanine bases on one strand hydrogen bond only with cytosine bases on the other strand (three hydrogen bonds). This specific interaction of bases is known as **base pairing**, and the two bases involved are called a base pair (Figure B.72). Each base pair consists of one purine base and one pyrimidine base.

The structure of DNA has been likened to a ladder which has been twisted into a helix. The base pairs are the rungs of the ladder, whereas the sugar–phosphate backbones are the sides of the ladder (Figure B.71).

The sequence of bases in one strand of DNA can be worked out by knowing the sequence in the complementary strand and remembering the rules:

A pairs with T
C pairs with G
In RNA, U pairs with A

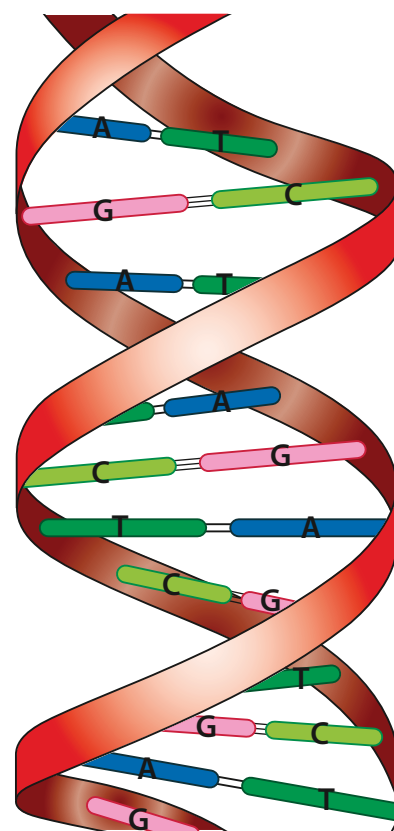
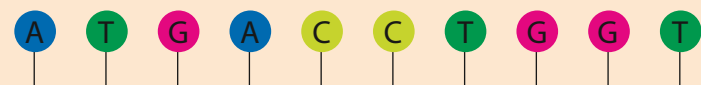


Figure B.71 The double-helix structure of DNA.

Worked example

B.12 The diagram below shows the sequence of bases in one strand of DNA. Work out the sequence of bases in the complementary strand.



We know that A pairs with T, and that C pairs with G. Therefore, from left to right, the sequence of bases in the complementary strand is T–A–C–T–G–G–A–C–C–A.

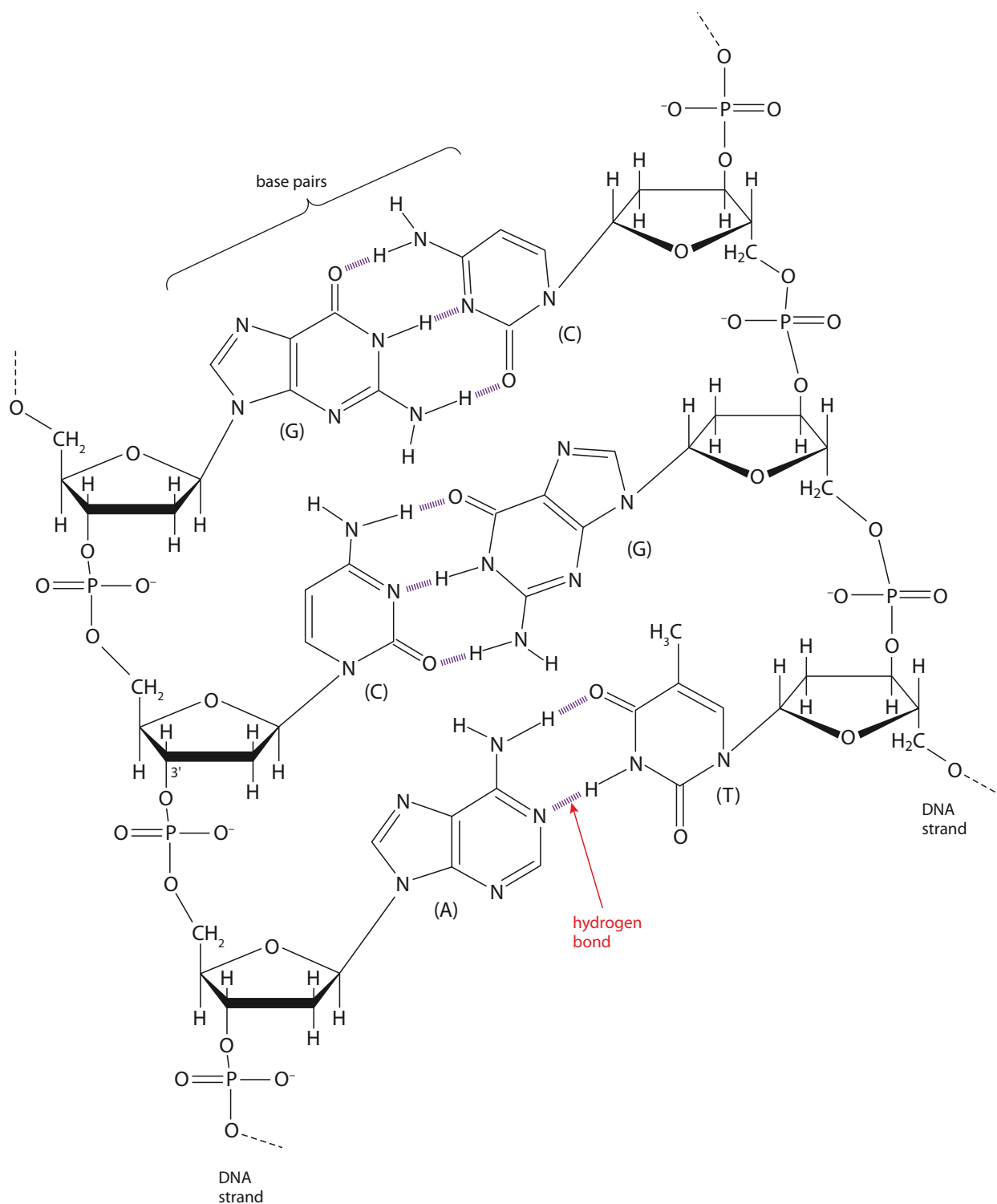


Figure B.72 Complementary base pairing in DNA.

DNA carries the genetic code

The nucleus of the cell houses the **chromosomes** – in humans there are 46 chromosomes per nucleus, made up of two sets of 23 chromosomes (one set inherited from each parent). Chromosomes are long, coiled strands of DNA wrapped around proteins called **histones** – these proteins are positively charged (because of charges on side groups) and are attracted to the negatively charged phosphate groups in the DNA. The DNA in these chromosomes contains **genes**, which carry all the information needed for an individual to grow, live and reproduce.

Genes are stretches of nucleotides in the DNA. Each gene has a specific sequence of nucleotides. As the sugars and the phosphates within DNA are fixed, it is the sequence of nucleotide bases that gives each gene its originality.

DNA replication

When a cell divides, it needs to copy its DNA so that the new cell will have an identical set of chromosomes to the parent cell. The process of copying DNA is called **DNA replication** and involves unwinding and separating the two strands of DNA by breaking the hydrogen bonds between the base pairs. This exposes the nucleotide bases of the DNA (which were sited in the interior of the helix) and allows the bases on each of the DNA strands to be used as a template for the formation of a **complementary** DNA strand (Figure B.73).

The formation of the new strands occurs by free nucleotides forming complementary base pairs with the DNA template strands – guanine nucleotides pair with cytosine bases on the DNA strand; thymine nucleotides pair with adenine bases on the DNA strand, and so on. The free nucleotides are joined together by phosphodiester links to form the new strands of DNA.

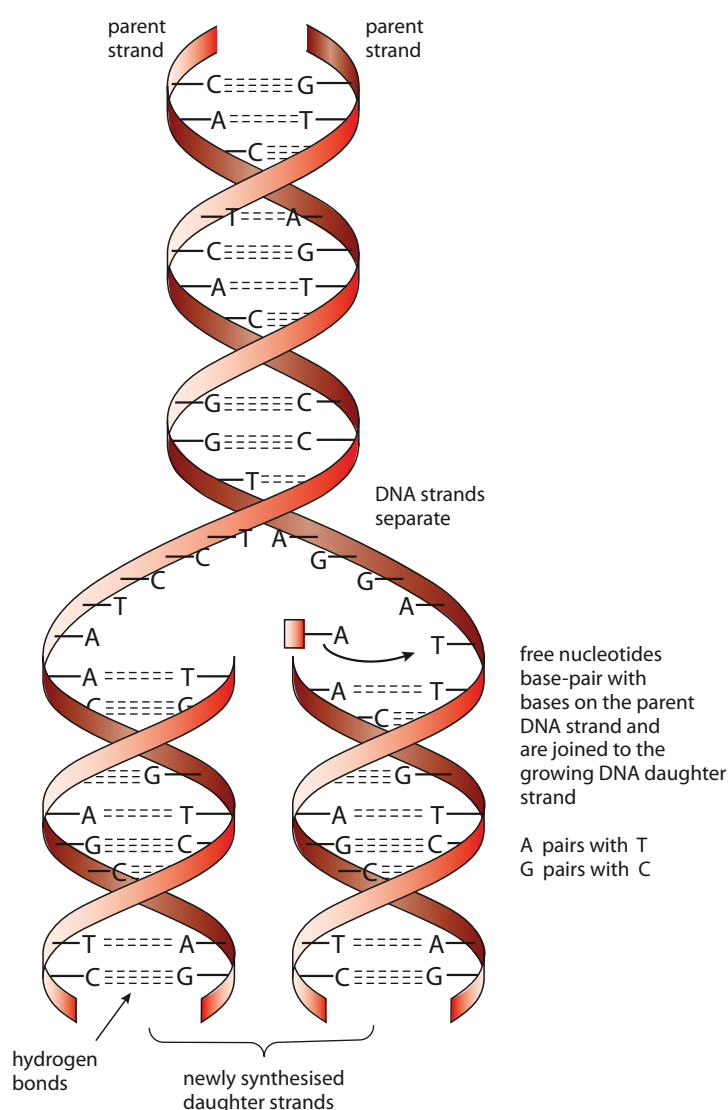


Figure B.73 DNA replication.

DNA transcription

To understand how genes carry the code for life, we need to know that each gene carries the code for the production of a single protein. It is these proteins, produced by the cell, that carry out the thousands of biochemical processes responsible for life.

Proteins are made up of a specific sequence of amino acids (called the primary structure). This sequence gives the protein its shape and therefore its function, and it is the sequence of nucleotide bases within each gene that dictates the primary structure of the protein produced. In other words, the information for the structure of each protein is carried in the sequence of nucleotide bases within the particular gene.

For a cell to produce a protein, the gene must first undergo a process called **transcription**. This occurs in the nucleus, and the first step in this process is the unwinding and separation of the two strands of DNA on which the gene is situated. In a similar manner to DNA replication, this exposes the nucleotide bases of the gene and allows the bases on one of the DNA strands to be used as a template.

Transcription differs to DNA replication, however, in that only one strand of DNA is used as a template and the complementary strand produced is a **ribonucleic acid** called **messenger RNA (mRNA)**. The complementary strand of mRNA is built up through complementary RNA nucleotides (called ribonucleotides) forming base pairs with the exposed bases of the DNA template. Guanine pairs with cytosine and adenine pairs with uracil (not thymine, as in DNA). As each ribonucleotide comes in and forms a base pair, it is joined to the growing mRNA chain by a phosphodiester link. This results in the production of a strand of mRNA that has the complementary sequence of bases to the gene of the DNA template strand. This mRNA then leaves the nucleus and enters the **cytoplasm**, where it takes part in the second process needed to produce a protein, known as **translation**.

Translation of the genetic code

Translation is the process of protein synthesis in which the code held in the sequence of bases of the mRNA is translated into the sequence of amino acids (primary structure) of the protein.

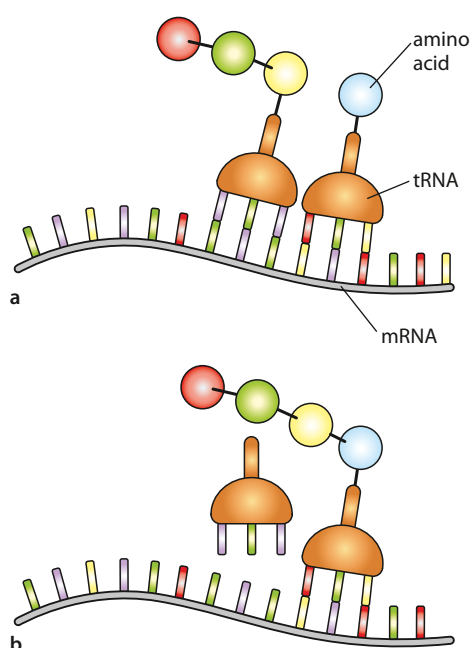


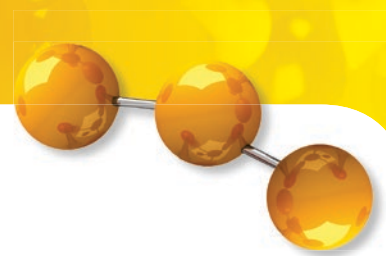
Figure B.74 The process of translation.

a Two tRNA molecules interacting with the codons on the mRNA. When the second tRNA molecule (carrying the blue amino acid) comes in and interacts with the mRNA, the growing polypeptide chain is cleaved from the first tRNA molecule and joined to the amino acid on the second tRNA molecule.

b The first tRNA molecule then leaves and a new tRNA molecule comes in and interacts with the next codon on the mRNA chain. The polypeptide chain will then be joined to the amino acid on this new tRNA molecule, and so on ... and so on, until the complete polypeptide chain has been produced.

During translation (Figure B.74), the mRNA first attaches to a ribosome (a small organelle in the cytoplasm) and the code is read by a type of RNA called **transfer RNA (tRNA)**. tRNAs are small RNA molecules with an amino acid covalently attached to one region of the molecule. Another region of the same tRNA molecule interacts with complementary bases on the mRNA. The tRNA interacts with a sequence of **three nucleotide bases** on the mRNA – these three bases are called a triplet code, or **codon**, and the three complementary bases on the tRNA are called an **anticodon**.

Each codon corresponds to only one amino acid (although several codons may correspond to the same amino acid there are 64 possible



codons but only 20 amino acids). For example, the codon UUU on mRNA corresponds to the amino acid phenylalanine. By this, we mean that a tRNA molecule with the anticodon AAA (remember A base pairs with U) will have a phenylalanine amino acid attached to it. The anticodon of tRNA will interact with the complementary codon on the mRNA and the phenylalanine amino acid will get incorporated into the growing polypeptide chain. The mRNA is read sequentially, so the next three bases (codon) will be exposed in the ribosome and then the complementary tRNA molecule bearing its specific amino acid will interact with it, incorporating the amino acid into the polypeptide chain, and so on, and so on (Figure B.74). Thus the sequence of bases in the DNA dictates the sequence of amino acids in the protein synthesised.

The **genetic code** refers to how the four-base code in DNA determines the sequence of 20 amino acids in proteins. Each three base codon codes for only one amino acid and this code is the same in all organisms – it is universal.

Genetically modified organisms

Genetically modified organisms (GMO) have genetic material that has been changed in some way by genetic engineering.

Genetic modification usually involves the insertion or deletion of specific genes in the GMO to produce desirable characteristics. In the production of **transgenic** species, genes are inserted from one organism into a completely unrelated one – this is partly where the controversy lies.

A GM food is one derived or produced from a GMO. The food can be substantially different from or essentially the same as the conventional food.

GM foods have been available in the marketplace since the 1990s. Most GM foods are plant-based and include products such as vegetable oils, fruits, vegetables and rice. Some animal-derived GM foods have been developed including (in 2006) a transgenic pig genetically modified to be rich in ‘good’ omega-3 fatty acids.

In the transgenic pig, the inserted gene, designed to make the pig produce more omega-3 oils, came from the microscopic nematode worm, *Caenorhabditis elegans*.

Benefits of GM foods

Much maligned, GM foods also have potential benefits.

- 1 Increased nutritional content of crops – for example, ‘golden rice’ has been genetically modified by inserting genes from a daffodil and a bacterium so that it is enriched with provitamin A. This has been suggested as a solution for vitamin A deficiency in certain parts of the world.
- 2 Increased yield of crops – of particular benefit to developing countries with exploding populations. Increased yield also permits diversification in the uses of some crops (e.g. use of vegetable oils as biofuel – a renewable alternative to gasoline – or corn to make biodegradable plastics).
- 3 The ability of crops to grow in adverse conditions – for example, the ability to grow in a hot climate with limited water and mineral availability. It might be possible to grow GM plants that are able to

remove pollutants (such as heavy metals) from barren soil, allowing it to be farmed again.

- 4 Improved crop resistance to disease, drought, pests and herbicides – inserted genes could confer these beneficial characteristics on crops to prevent damage from microbial infection or attack by damaging pests such as beetles and locusts. For instance, Bt corn contains a gene from a soil bacterium which produces a protein that kills the European corn borer, an insect responsible for millions of dollars' worth of damage to corn crops each year. Using this strain of corn means that less insecticide has to be used, which saves money and is also good for other non-harmful insects in corn fields.
- 5 Enhanced taste, texture, quality and shelf life – improves on currently available foods and makes GM products more desirable and saleable.
- 6 Production of 'healthy' crops – which can produce higher levels of useful unsaturated fatty acids and fewer saturated ones. Organisms can also be modified to reduce allergic reactions – for instance a GM cow has been bred that produces milk which is less likely to cause allergic reactions.
- 7 Improved animal health – GM animals could also be developed to be more resistant to disease.
- 8 Improved conservation of water, soil and energy – GM crops can be engineered to require less space, less water and less management.

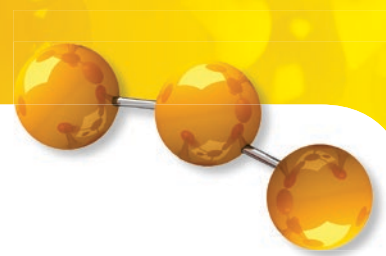


Standards of food labelling vary from country to country. How much information do we need to make an informed choice? Can too much information be bad? If every food was labelled with detailed medical information concerning the benefits and adverse effect of the components, environmental effects of the production process etc. would we be better informed or just more confused? How do we decide what information is essential and what is not?

Concerns about GM foods

There is significant lobbying and opposition to the use and globalisation of GM foods. Some of the arguments against genetic modification include:

- 1 Not enough is known about the long-term effects – GM foods are a fairly modern invention, so not enough is known about the long-term effects of their introduction into the environment. Effects on the environment and human health may take many more years to be apparent.
- 2 Escape of transgenic material into the 'wild' – possible contamination of indigenous, natural species with unknown consequences. Imagine the effects of a disease- and herbicide-resistant gene becoming fully integrated into several species of pervasive weeds. A gene conferring antibiotic resistance could possibly be transferred to a species of bacteria making it more difficult to treat infection. There are worries about genetically modified DNA being transferred to the gut bacteria of people eating GM foods.



- 3 Allergies – there are worries that GM foods could cause more and/or different allergies.
- 4 Damage to the environment and unbalancing of ecosystems may lead to the endangerment of multiple species and upset the current balance of ecosystems and food chains worldwide – for instance, there are worries that some of the byproducts of Bt corn could kill other insects and disrupt ecosystems.
- 5 Alteration of the composition of a balanced diet – by changing the nutritional content and quality of foods.
- 6 Exploitation and monopolisation – by countries and companies with intellectual property rights to the technology. These GM foods are generally developed by large biotechnology companies, and the research and development is expensive so they have to get the money back somehow. These companies control the distribution and price of seeds for GM foods. Farmers in developing countries may become dependent on GM seeds controlled by multinational companies.
- 7 Ethical considerations – such as tampering with nature, ‘playing God’ and general disgust at consuming GM plant and animal products. There are also concerns about labelling in that it is not always clear whether food contains GM ingredients or not.

Nature of science

Many scientists over many years contributed to the discovery of the structure of DNA and their investigations involved different methods. Careful chemical analysis, building physical models and finally X-ray crystallography all played their role in the development of the understanding of the structure. The development of technology played a major role in this process as the double helix structure of DNA was finally proposed after examining X-ray diffraction data.



Test yourself

- 30 Work out the nucleotide sequence in the complementary strand of DNA for each of the following fragments:
 - a T-T-C-G-G-A-C
 - b A-A-A-C-G-C-C-T-A-T-T-T-G-A-C-C
- 31 Work out the nucleotide sequence in the complementary strand of RNA for each of the following fragments:
 - a U-C-A-A-G-U
 - b G-G-G-U-U-A-C-U-G-C-C-A

Learning objectives

- Understand that biological pigments are coloured compounds produced by metabolism
- Understand, in terms of conjugation/delocalisation, why pigments absorb visible light
- Understand the structures and properties of anthocyanins and carotenoids
- Understand the structure of porphyrin complexes
- Explain the shape of hemoglobin's oxygen dissociation curve
- Understand the factors that affect the affinity of hemoglobin for oxygen
- Understand that carbon monoxide is a competitive inhibitor for oxygen-binding in hemoglobin
- Describe the greater affinity of fetal hemoglobin for oxygen than adult hemoglobin
- Understand that cytochromes contain heme groups and are involved in redox reactions in cells
- Describe the function of pigments in trapping light energy in photosynthesis
- Understand that pigments can be investigated using paper and thin-layer chromatography

B9 Biological pigments (HL)

Biological pigments are coloured compounds produced by metabolism – they are made by chemical reactions in cells.

Biological pigments include anthocyanins, carotenoids, chlorophyll and heme.

Chromophores

In order to absorb electromagnetic radiation in the UV–Vis region of the spectrum, molecules must generally contain a double bond in the form of $C=C$, $C=O$ or a benzene ring. These groups, which give rise to absorptions in the UV–Vis region, are called **chromophores**. Electromagnetic radiation in the ultraviolet–visible region of the spectrum is absorbed to promote electrons from a low energy level (molecular orbital) in molecules to a higher energy level (molecular orbital).

Conjugated systems

A **conjugated system** is a sequence of alternating single and double bonds in a molecule.

The bonds highlighted in Figure B.75 form a conjugated system, but the two double bonds at the ends of the molecule are not part of this system because they are separated from the other double bonds by more than one single bond.

The double bonds must alternate with single bonds for a system to be conjugated – if there are two or more single bonds between the double bonds then the system is not conjugated.

We have already seen that a benzene ring can be represented as a ring with three alternating double and single bonds – so a benzene ring is a conjugated system. The molecule shown in Figure B.76 has six conjugated double bonds.

Electrons are delocalised in a conjugated system.

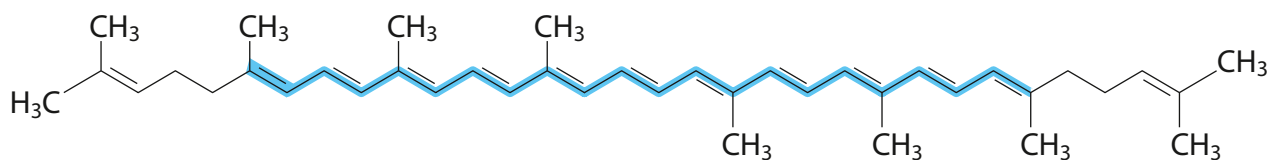


Figure B.75 Lycopene, the red pigment in tomatoes, has 11 conjugated double bonds. Note that not all the $C=C$ bonds in lycopene are part of the conjugated system.

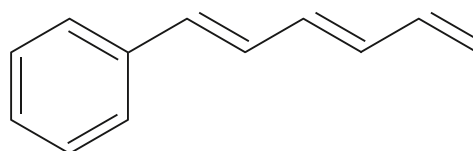


Figure B.76 A molecule with six conjugated double bonds.

We can see from Figure B.77 that electrons are delocalised in a conjugated system because p orbitals can overlap along the whole chain.

Absorption of electromagnetic radiation and colour

For a compound to be coloured, its molecules must absorb visible light. Visible light is electromagnetic radiation with wavelengths between about 400 and 750 nm. Therefore if a molecule absorbs radiation between these wavelengths it will be coloured. The longer the conjugated system, the longer the wavelength of the radiation absorbed and if a conjugated system involves more than about eight double bonds, the molecules should absorb in the visible region of the spectrum and be coloured.

The longer a conjugated chain (delocalised system), the longer the wavelength of radiation absorbed by a molecule.

Lycopene, which has a system of 11 conjugated double bonds (Figure B.75), absorbs light in the blue–green part of the visible spectrum and therefore appears red. Retinol (Figure B.78), however, only has a system of five conjugated double bonds and therefore does not absorb visible light (it only absorbs ultraviolet radiation) and is colourless.

Chlorophyll a and b have long conjugated systems (highlighted in Figure B.79). They absorb light in the 400–500 nm region and in the 600–700 nm region (Figure B.80). The green light in the middle of the spectrum is not absorbed, and so these molecules look green in natural light.

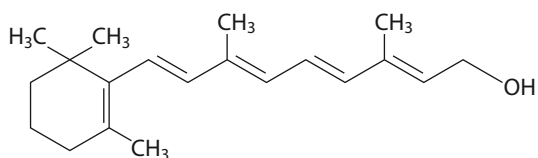


Figure B.78 Retinol has only five conjugated double bonds and absorbs only in the UV region of the electromagnetic spectrum.

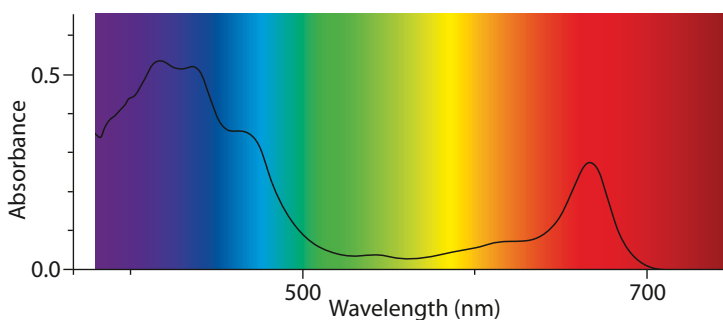


Figure B.80 The visible spectrum of chlorophyll.

Nature of science

Quantitative data is very important in science and collection of data concerning the wavelength absorbed and the absorbance of a solution provides a lot more information than descriptions just based on colour.

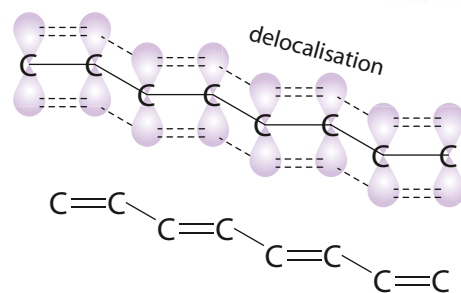


Figure B.77 Delocalised electrons and overlapping p orbitals in a conjugated system.

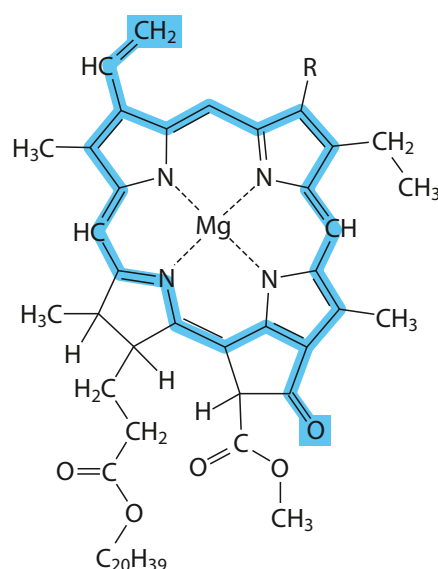


Figure B.79 The basic structures of chlorophyll a (R is CH_3) and chlorophyll b (R is CHO), showing the conjugated system. Chlorophyll b has an extra double bond that is part of the conjugated system.

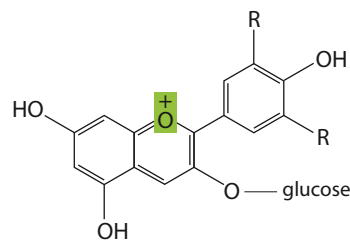


Figure B.81 The basic anthocyanin structure. The conjugated/delocalised system is highlighted in green (it could also be extended to include lone pairs on O atoms). Other sugars may also be present instead of glucose.

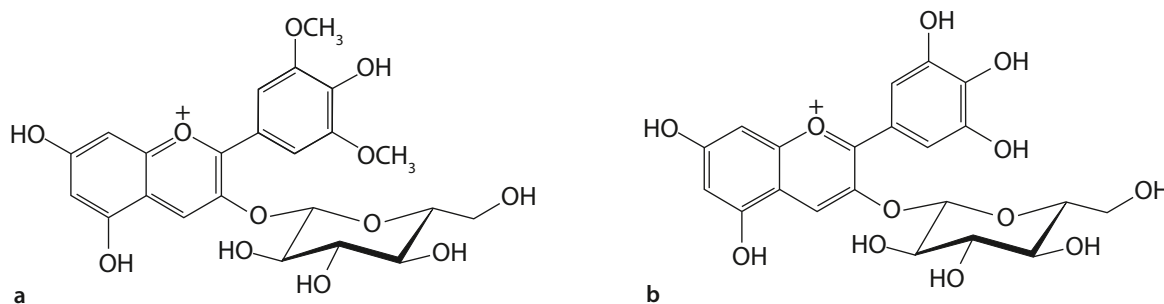


Figure B.82 **a** Oenin found in purple grape skin; **b** myrtillin found in blackcurrants and blueberries.

Anthocyanins

Anthocyanins are very common pigments in plants. They all have a characteristic structure based on the same core unit (Figure B.81).

Anthocyanins are the principal pigments responsible for the pink, red, blue and purple colours of many fruits and vegetables including red cabbage, blackcurrants, strawberries, cranberries, blueberries, raspberries and grapes (Figure B.82).

Anthocyanins have molecules that contain aromatic rings but are soluble in water because they also have a large number of OH groups which can hydrogen bond to water.

The presence of metal ions can affect the colour of anthocyanins.

Anthocyanins can form vivid, deep-coloured complexes with metal ions such as Al^{3+} and Fe^{3+} . Complex ions are formed with the anthocyanin molecules acting as ligands.

Carotenoids

Carotenoids are the most widespread pigment found in nature – this is primarily due to their abundance in algae. Carotenoids generally absorb in the blue–violet region of the visible spectrum and therefore transmit or reflect longer wavelengths of the visible spectrum and so have colours in the yellow–orange–red region. Carotenoids are present in carrots, tomatoes, watermelon, sweet peppers and saffron etc.

Most carotenoids are derived from a 40-carbon polyene chain (multiple $\text{C}=\text{C}$ double bonds). The ends of the chain may terminate in a cyclic (ring) group which may or may not have oxygen-containing functional groups attached.

Those carotenoids that contain solely carbon and hydrogen are called **carotenes**; those containing oxygen atoms are called **xanthophylls**.

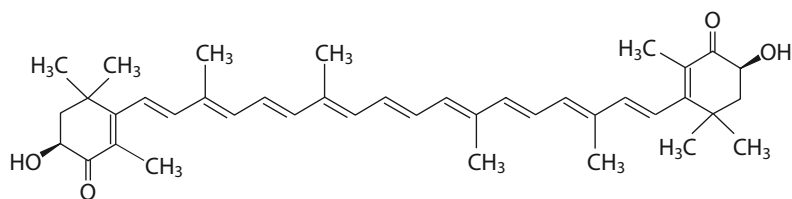


Figure B.83 The red carotenoid astaxanthin gives live lobsters and crabs their blue–green hue when it is complexed with protein. When the shellfish is cooked, the astaxanthin dissociates from the protein and the shell turns red. Astaxanthin is also responsible for the deep-pink colour of wild salmon.

Lycopene (Figure B.75), β -carotene and astaxanthin (Figure B.83) are carotenoids. They consist of mainly non-polar hydrocarbon chains and tend to be insoluble in water but soluble in lipids (fats).

Carotenoids are involved in light-harvesting in plants during photosynthesis – a photon of light is absorbed by a carotenoid molecule to promote an electron to an excited state. The energy it has absorbed is then transferred to chlorophyll.

Stability of pigment colours

A number of factors affect the stability of pigments – variation in these may result in a loss or change of pigment colour.

Major factors affecting pigment colour are temperature, pH, oxidation, the presence of metal ions and the oxidation number of metal ions. These factors contribute to changes in the structure of the pigment and/or the way in which it absorbs certain wavelengths of light.

Effect of pH on anthocyanins

In aqueous solution, anthocyanins exist in a complex equilibrium between four different structural forms:

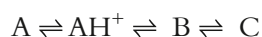


Table B.3 summarises the different structures and colours.

Anthocyanin	Structural form	Colour
A	quinoidal base	purple/blue
AH^+	flavylium	red
B	carbinol base	colourless
C	chalcone	yellow

Table B.3 Structural forms and colour of anthocyanins.

Which species is most stable, and hence the position of the equilibrium, depends on the pH and the temperature of the solution. At low pH the red flavylium form predominates (Figure B.84). This is converted to the carbinol base form as the pH is increased. This has a shorter conjugated system than the flavylium cation so that it absorbs electromagnetic radiation in only the UV region of the spectrum and is therefore colourless. This species dominates at pH 4–5 and the colour of the mixture in this pH range will be quite pale. As the pH is increased the carbinol base form is converted to the yellow chalcone form. When the pH is increased further, the purple quinoidal base form is also formed which is in equilibrium with its intensely blue-coloured anion.

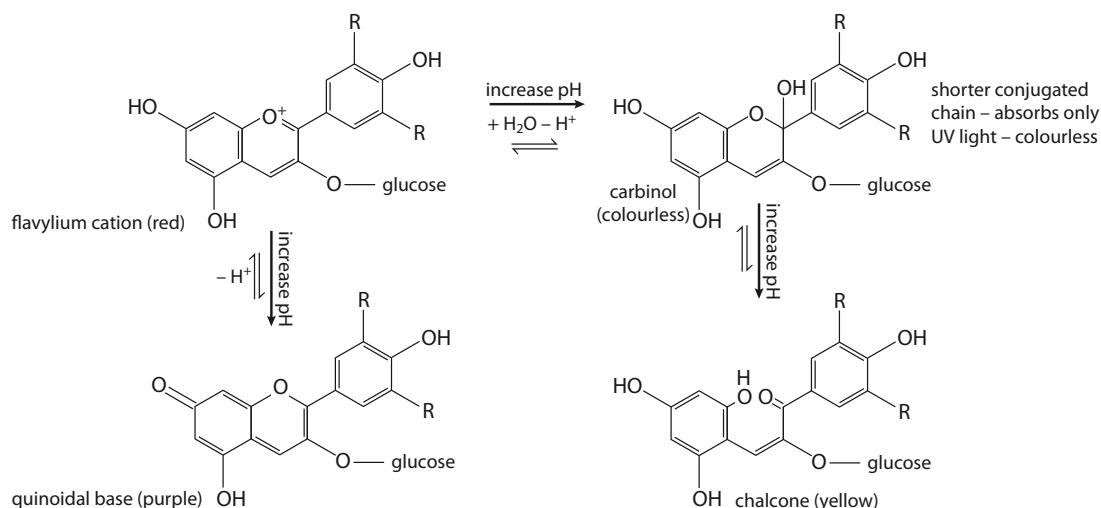


Figure B.84 A complex equilibrium, which is very sensitive to pH, exists in an anthocyanin solution.

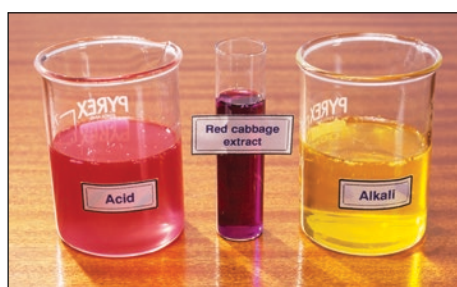
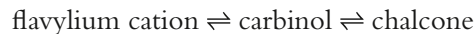


Figure B.85 Anthocyanins extracted from red cabbage can be used as acid–base indicators.

Because the colour changes with pH, anthocyanins can be used as indicators (Figure B.85).

Effect of temperature on anthocyanins

The flavylium cation – the form that is most important for the colour of species containing anthocyanins – is less stable at higher temperatures. At low temperatures (and low pH), the red form is abundant but as the temperature rises the equilibria:



shift to the right causing the pigments to lose their red colour.

Anthocyanins also dissociate into smaller molecules that do not absorb visible light as the temperature is increased. The presence of oxygen accelerates the rate of thermal degradation.

Effect of different conditions on carotenoids

Due to the presence of a conjugated multiple C=C double-bond system, carotenoids such as β -carotene are not only coloured but also highly susceptible to oxidation catalysed by light (photo-oxidation) and oxidation by metal ions and organic hydroperoxides (R–O–O–H). On oxidation, the conjugation is destroyed and the pigment is bleached (decolorised). The oxidised species cannot be converted to vitamin A and often have unpleasant odours.

Carotenoids are thermally stable up to 50 °C and structurally stable in acidic conditions (pH range 2–7). However, when heated above 50 °C the naturally occurring all-*trans* form rearranges to yield a variety of *cis* isomers. This isomerisation is also influenced by the presence of light and chemicals such as iodine.

Effect of different conditions on chlorophylls

Chlorophylls can be destabilised by high temperatures depending on the pH of the environment. Chlorophyll remains stable in alkaline solution

(pH 9), but in an acidic solution (pH 3) it is highly unstable. When plant material is heated – as in cooking – the plant cell membranes break down and cause acid to be released, lowering the pH of the surrounding solution. This causes vegetables to lose their green colour. The magnesium ion located at the centre of the porphyrin ring (see Figure B.79) is normally stable and difficult to remove but at low pH the Mg^{2+} ion is displaced by two H^+ ions, resulting in formation of an olive-brown **pheophytin complex**.

Generally, brighter-green vegetables are viewed as being more appealing than darker-coloured ones. By reducing cooking time and boiling vegetables with the pan lid off to allow the escape of volatile acids, the production of pheophytin and discoloration can be minimised. Adding a small amount of sodium hydrogencarbonate (NaHCO_3) to the water during boiling may also help to keep the vegetables green because this raises the pH of the cooking solution.

The heat-induced cellular degradation caused by cooking can also make chlorophyll pigments more likely to undergo photodegradation, in which they chemically decompose in the presence of light.

Porphyrin rings (heme and chlorophyll)

The structures of heme and chlorophyll are shown in Figure B.86.

The principal structure in both heme and chlorophyll (Figure B.87) is based on a complex, planar **macrocyclic** unit called a **porphin ring** which contains a system of conjugated $\text{C}=\text{C}$ double bonds. A porphin ring with side groups attached at positions 1–8 is called a **porphyrin** and both heme and chlorophyll are described as porphyrin complexes.

Chlorophylls contain a porphyrin unit complexed to a central Mg^{2+} ion. Heme is a complex between a porphyrin unit and an Fe^{2+} ion.

Heme acts as a **prosthetic group** in both myoglobin (the pigment in muscles) and hemoglobin (the pigment in red blood cells). In both a heme group is associated with a polypeptide chain in a 1 : 1 ratio.

The term ‘macrocycle’ is used in many ways – in this context it means a large ring with multiple donor atoms (nitrogen atoms) that can bond to metal ions.

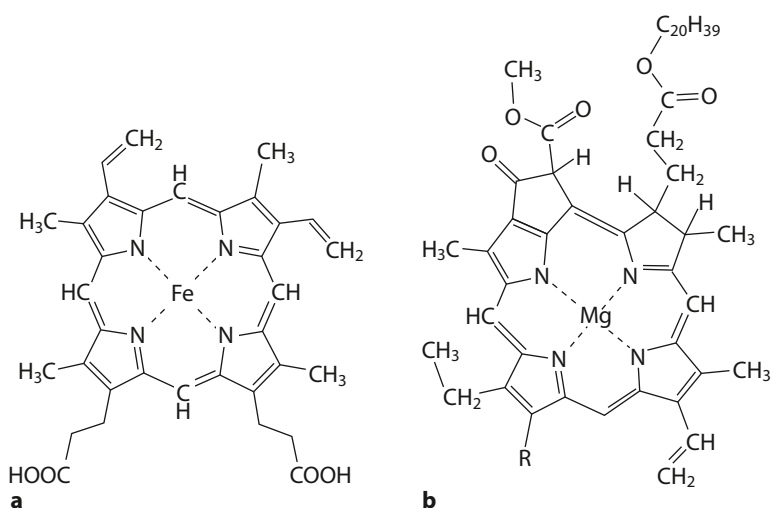


Figure B.86 The structures of **a** heme and **b** chlorophyll. The most commonly met form of heme is heme B (illustrated).

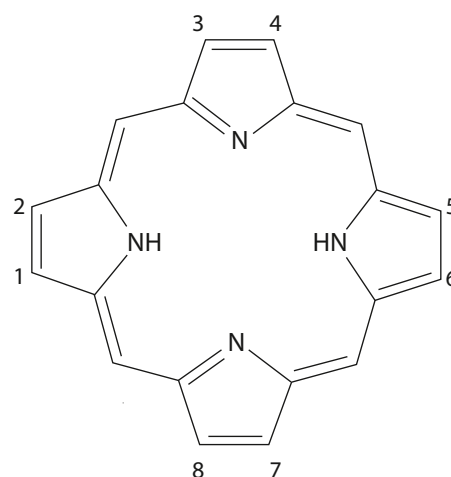


Figure B.87 A porphin ring – a macrocycle.

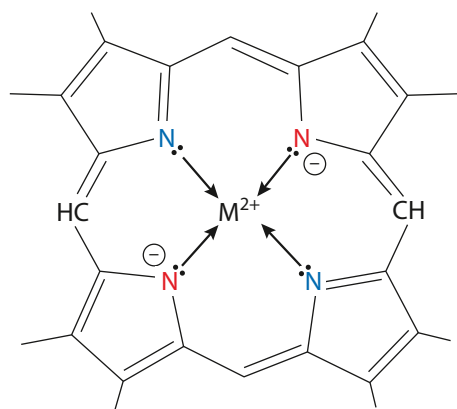


Figure B.88 The bonding in porphyrin complexes.

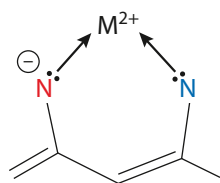


Figure B.89 One of the rings formed by coordination of a ligand to a metal ion.

The bonding in porphyrin complexes

We have already met the idea that transition metals can form complexes with ligands such as water and ammonia in Topic 3. Ligands that bind to the transition metal through only one atom are called monodentate ligands (sometimes called unidentate ligands). However, there are ligands that can bind through more than one atom and these are called polydentate ligands. The porphyrin ring is a **tetradentate** ligand because it coordinates to the central ion through four nitrogen atoms.

Part of the structure of a porphyrin complex is shown in Figure B.88. The two nitrogen atoms in red form only two bonds to the rest of the ligand and so must be negatively charged to have a full outer shell (they have lost H^+ ions – compare with Figure B.87). The M ion at the centre has a 2+ charge. The bonding can be regarded as involving coordinate (dative) covalent bonds between the lone pairs on the nitrogens and the M^{2+} ion.

Polydentate ligands are also called chelating ligands and they form chelate complexes (often just called chelates) with transition metal ions. These complexes contain a ring which includes the metal ion (Figure B.89).

The binding of oxygen to hemoglobin

Hemoglobin transports oxygen from the lungs through the bloodstream and releases it to the cells of the tissues to carry out respiration.

Hemoglobin consists of four polypeptide sub-units, each of which contains a heme prosthetic group with the iron at the centre of the heme having oxidation number +2. Each heme can carry one molecule of oxygen, so each hemoglobin unit can transport four molecules of oxygen.

The iron in the heme can bond to six ligands. In the unbound state, the Fe^{2+} is bonded to five ligands – four are the nitrogen atoms of the porphyrin and the other is an amino acid that attaches it to the protein. When molecular oxygen binds, this becomes the sixth ligand, and hemoglobin is said to be **oxygenated** (it is this oxygenated form that gives blood its red colour, because of the red colour of the heme prosthetic group). Binding of the oxygen molecules results in Fe^{2+} being oxidised to Fe^{3+} . In hemoglobin, the oxygen binds reversibly, allowing its release to tissue cells to be used in cellular respiration.

The graph in Figure B.90 shows how the affinity of hemoglobin for oxygen changes as the partial pressure of oxygen changes. The scale on the y-axis represents the fraction of iron ions bound to oxygen molecules. This is called an oxygen binding curve or oxygen dissociation curve.

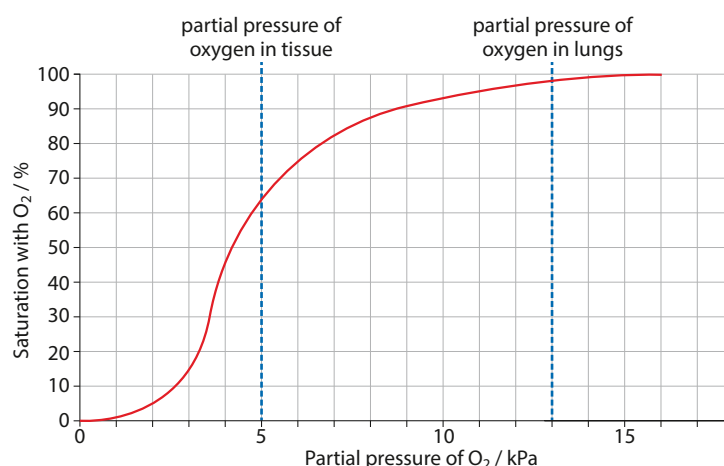


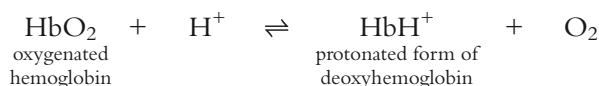
Figure B.90 The oxygen binding/dissociation curve for hemoglobin. The partial pressure of oxygen is the pressure of the oxygen in a mixture of gases.

The type of curve in Figure B.90 is described as sigmoidal. When the partial pressure of oxygen is low, hemoglobin has a low affinity for oxygen but the affinity increases markedly as the partial pressure of oxygen increases – the gradient of the curve increases. This suggests that it becomes easier for oxygen to bind to hemoglobin when some oxygen molecules have already bound to the iron – and that the binding of oxygen is cooperative. Hemoglobin has a tetrameric structure with four iron–heme complexes. The binding of oxygen to one of the iron ions in the tetramer changes the shape (conformation) of the protein in such a way that it becomes easier for oxygen molecules to bind to the other sites. This is an allosteric effect (see page 51) – the binding of a molecule at one site has an effect on another site.

This type of curve is important for the functioning of hemoglobin as an oxygen carrier – the affinity for oxygen is high when the blood passes through the lungs (high partial pressure of oxygen) so the hemoglobin binds lots of oxygen, but the affinity is much lower in the tissues and so the hemoglobin gives up the oxygen to the tissue (because it is bound to more oxygen than it can be at that partial pressure).

The effect of pH, carbon dioxide and temperature on the binding of oxygen by hemoglobin

Hemoglobin also transports H^+ ions and CO_2 molecules around the body. Both pH and concentration (partial pressure) of carbon dioxide affect the ability of hemoglobin to bind oxygen. As pH decreases ($[\text{H}^+]$ increases) the affinity of hemoglobin for oxygen decreases. This can be represented by the equilibrium:



As the H^+ concentration increases, the position of this equilibrium shifts to the right and O_2 is released from the hemoglobin. The H^+ does not bind to the same site as the O_2 but rather an amino-acid side-chain. Binding H^+ changes the shape (conformation) of the protein slightly to reduce the affinity for oxygen.

A higher carbon dioxide concentration also reduces the affinity of hemoglobin for oxygen. Carbon dioxide produced in respiration in cells diffuses into red blood cells, where it dissolves to form an acidic solution (carbonic acid, H_2CO_3). This lowers the pH so that more H^+ binds to hemoglobin and causes a release of oxygen. Hemoglobin also binds carbon dioxide, but not at the same site as O_2 . This carbon dioxide reacts with the NH_2 group on the end amino acid of each polypeptide chain that makes up hemoglobin (Figure B.91). This process has two effects – it releases H^+ and changes the shape of the protein; both of these reduce the affinity of hemoglobin for oxygen.

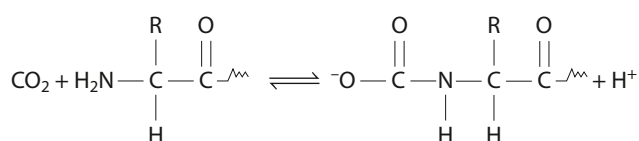
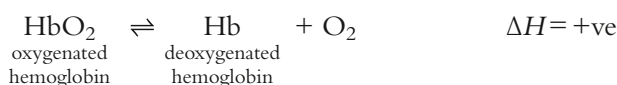


Figure B.91 The reaction of carbon dioxide with a terminal amino acid of a polypeptide chain.

The ability of hemoglobin to bind oxygen decreases as the temperature increases. This suggests that the oxygenation process is exothermic and that the deoxygenation process is endothermic:



Increasing the temperature causes the position of equilibrium to shift to the right – towards the deoxygenated form, causing the release of oxygen.

The affinity of iron in hemoglobin for carbon monoxide is approximately 200 times the affinity for oxygen.

Carbon monoxide also has another effect on the ability of hemoglobin to deliver oxygen to cells – it reduces the ability of hemoglobin to release oxygen.

The effect of carbon monoxide

CO is a better ligand than O₂ and binds to the iron ions in hemoglobin more strongly than oxygen does, so carbon monoxide is a competitive inhibitor for the binding of oxygen to hemoglobin. The lone pair on the carbon atom in carbon monoxide binds to the iron ion.

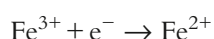
This can be compared with the effect of CO₂ and H⁺ which was non-competitive because they did not bind at the same site as oxygen.

Fetal hemoglobin

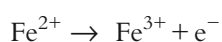
Fetuses (embryos) have a different type of hemoglobin (hemoglobin F) to adult humans (hemoglobin A). Fetal hemoglobin has a higher affinity for oxygen under the same conditions. This is important because it allows the transfer of oxygen from a mother's hemoglobin to that of her fetus.

Cytochromes

Cytochromes are proteins that absorb strongly in the visible region of the spectrum because of the presence of heme groups. Cytochromes are involved in key redox reactions in cells that result in the production of energy. In these reactions, the oxidation number of iron changes between +2 and +3. For instance, if a cytochrome accepts an electron from a reducing agent, Fe³⁺ is converted to Fe²⁺:



When this electron is transferred to an oxidising agent the Fe²⁺ is oxidised back to Fe³⁺:



Photosynthesis

The main pigment responsible for harvesting light energy from the Sun and converting it to chemical energy in the process of photosynthesis is **chlorophyll**. During photosynthesis light is absorbed by chlorophyll to promote electrons to higher energy levels. These electrons are then passed on via an electron-transport chain (a series of redox reactions) to a low-energy-electron acceptor. In the process the energy of the excited electrons is converted to chemical energy.

Carotenoids are also involved in light-harvesting in plants for photosynthesis – a photon of light is absorbed by a carotenoid molecule to promote it to an excited state. The energy it has absorbed can then be transferred to chlorophyll. Carotenoids such as β-carotene absorb visible light of wavelengths that are different from chlorophyll and therefore increase the amount of energy that can be obtained from light.

Investigation of pigments by chromatography

Paper chromatography and thin-layer chromatography (TLC) can be used to separate and identify mixtures of pigments. Paper chromatography for the separation and identification of amino acids has already been covered on page 13.

TLC is a very similar technique to paper chromatography and the process is carried out in basically the same way, but instead of a piece of paper, a plate (a piece of plastic, glass or metal) coated in silica gel or alumina is used.

Separation here happens because of adsorption – the components of the mixture are either dissolved in the solvent (mobile phase) or adsorbed onto the stationary phase. The greater the tendency of a solute molecule to be adsorbed onto the stationary phase, the more slowly it moves along the plate.

Because pigments are coloured they will be visible on the paper/plate without the need for using a locating agent (ninhydrin was needed for amino acids).

TLC and paper chromatography may both be used for investigating pigments but TLC has many advantages over paper chromatography – for example, it is faster and gives better resolution (better separation of spots).

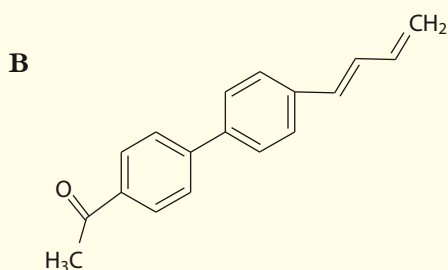
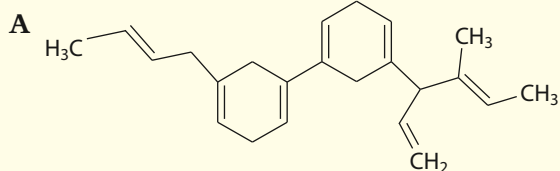
Adsorption means that molecules/ions ‘stick’ to the surface of the solid particles in the stationary phase.

Nature of science

The effective communication of ideas is essential if the general public is to gain a better understanding of science. In science, explanations at many different levels are possible and the explanations are usually tailored to suit the audience. For instance, the colours of substances can be explained most simply using the idea that chlorophyll is green because it absorbs red light and reflects/transmits green. We could go further and suggest that chlorophyll is coloured because it has a long conjugated system but the explanation goes further than this again ... why should conjugation mean that a substance absorbs visible light? The explanation eventually ends up with some quite complicated maths! It is impossible to get bored with science because you can always go deeper – rather like an onion, there is always another layer!

? Test yourself

- 32 State which of the following is more likely to be coloured.



- 33 State whether each of the following statements is true or false:

- a Chlorophyll is green because it absorbs green light.
- b The binding of one molecule of oxygen to hemoglobin make it easier for other oxygen molecules to bind.
- c Carbon dioxide is a competitive inhibitor for oxygen binding of hemoglobin.
- d Hemoglobin has a greater affinity for oxygen at pH 7.5 than at pH 7.0.
- e Fetal hemoglobin has a greater affinity for oxygen than adult hemoglobin.
- f Anthocyanins are fat/lipid-soluble.

Learning objectives

- Understand how the D and L notation can be used to assign the stereochemistry of a chiral carbon in sugar molecules
- Understand that naturally occurring amino acids all have the L form
- Understand the difference between α - and β -forms of sugars
- Describe the structure of cellulose and compare it with starch
- Understand that unsaturated fats usually have *cis* double bonds
- Describe the hydrogenation of fats
- Discuss the advantages and disadvantages of hydrogenation reactions
- Understand the role of retinal in vision

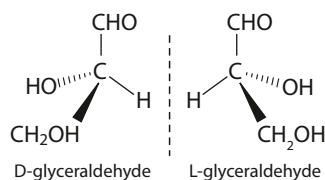


Figure B.92 D- and L-glyceraldehyde.

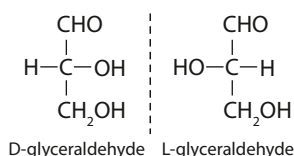


Figure B.94 Fischer projections for D- and L-glyceraldehyde.

B10 Stereochemistry in biomolecules (HL)

Stereoisomerism was introduced in Topic 10.

To exhibit optical isomerism, a molecule must have a chiral centre (a carbon atom with four different groups attached). Various methods are employed to describe the configuration of groups around the chiral centre.

The D and L system

The D and L convention for naming enantiomeric forms of a molecule tends to be used for carbohydrates and amino acids. It is rather an old-fashioned system and was developed before X-ray crystallography allowed the determination of the absolute configuration of molecules. When the D and L system is used, everything is compared to glyceraldehyde (2,3-dihydroxypropanal) – see Figure B.92.

Application of the D and L nomenclature for sugars is usually based on Fischer projections, which are a method of representing the structure of the straight-chain forms of sugars by projection on to a plane. In a Fischer projection, the sugar molecule is shown with the carbon numbered 1 at the top – according to the normal naming rules the aldehyde/ketone group will get the lowest possible number. Groups that point away from you are drawn vertically and those that point towards you are drawn horizontally (Figure B.93).

When applying the D and L system to larger sugar molecules it is the chiral carbon furthest away from the carbonyl group that is important.

When shown in Fischer projection, if the OH group on the chiral carbon furthest away from the carbonyl group is on the right, the molecule is assigned the label 'D'; if the OH group is on the left, it is given the label 'L' (Figure B.94).

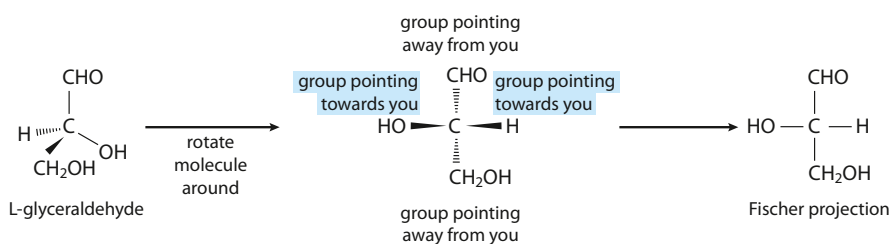


Figure B.93 How to derive a Fischer projection.

Looking at two forms of glucose in the Fischer projection, we can compare them to glyceraldehyde to determine which is the D enantiomer and which is the L form (Figure B.95). In D-glucose the OH group on the chiral centre furthest from the carbonyl group is on the right and in L-glucose it is on the left.

There are several chiral centres in glucose and in D- and L-glucose all chiral centres have the opposite configuration, so that the two molecules

are mirror images. If only some of the chiral centres have the opposite configurations, so that the molecules are not mirror images (Figure B.96), then the two molecules are diastereomers and are given different names – the configurations at the other chiral centres determines the name of the substance.

D sugars are much more common than L sugars in nature.

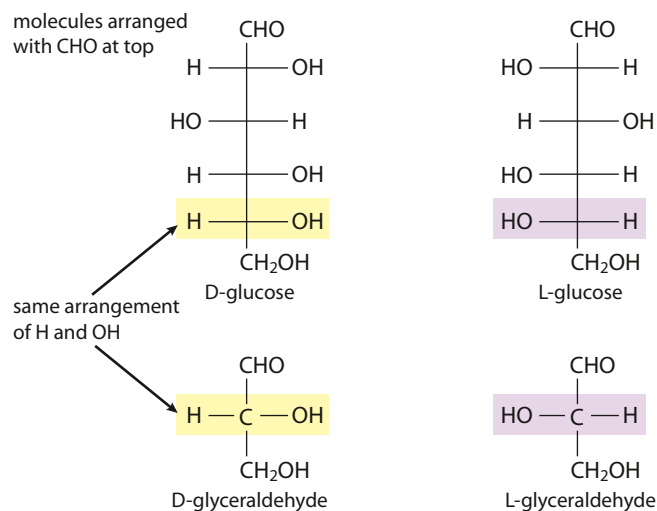


Figure B.95 Fischer projections of the two stereoisomers of glucose and their relationship to glyceraldehyde.

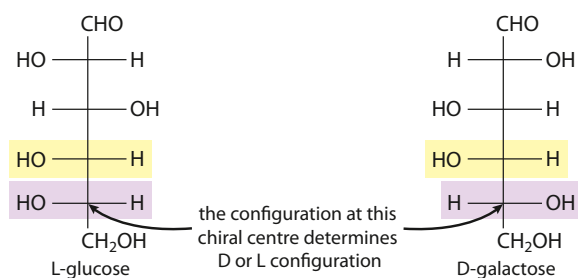
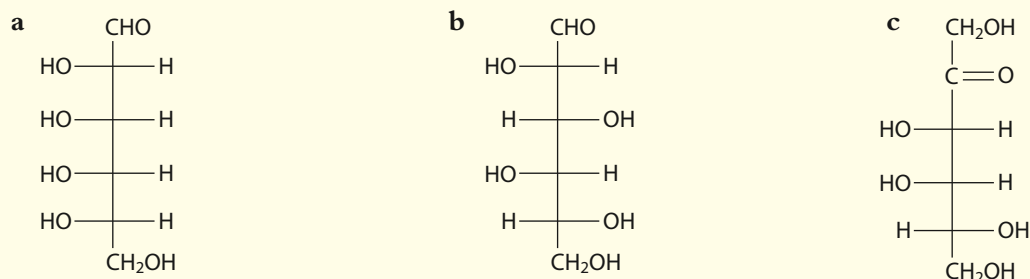


Figure B.96 L-glucose and D-galactose are diastereomers – they are not mirror images. The configuration at the chiral centre highlighted in yellow is the same for both.

? Test yourself

34 Work out whether each of the sugars shown below is the D form or the L form:



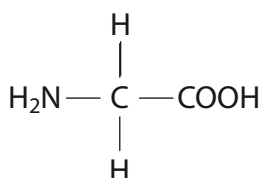


Figure B.98 Glycine does not have a chiral centre.

Amino acids

When applying the D and L system to amino acids, the ‘**CORN**’ rule is used. This looks at the positions of the **COOH**, **R** and **NH₂** groups around the chiral carbon.

The molecule is drawn in the classic tetrahedral form with the chiral carbon at the centre and the single hydrogen atom pointing away from you. If moving from the COOH to R to NH₂ you go in a clockwise direction, the structure is the D form; if you go anticlockwise it is the L form (Figure B.97).

All amino acids are chiral, except glycine which has only three different groups around the central carbon (Figure B.98).

All amino acids found in naturally occurring proteins are L-amino acids.

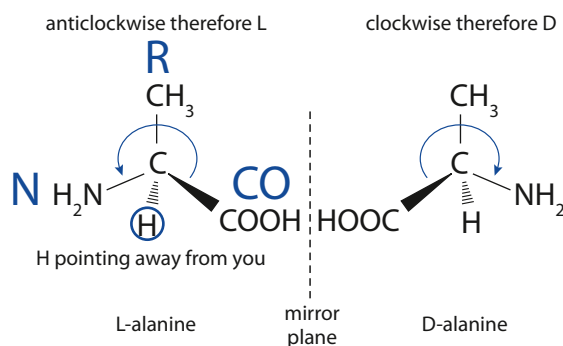
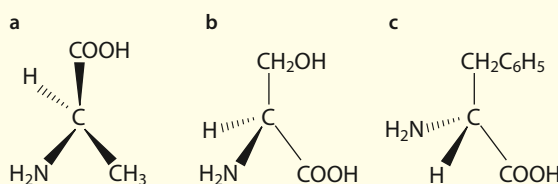


Figure B.97 D and L stereoisomers of the amino acid alanine.

? Test yourself

- 35 Classify each of the following amino acids as the D form or the L form.



α - and β -forms of the same sugar are called anomers.

Alpha and beta forms of sugars

When a sugar cyclises from the straight-chain form, an extra chiral centre is formed (Figure B.99). The oxygen on carbon 5 can either attack the C=O group from above the plane of the group or below and thus two possible cyclic molecules can be formed. These are called α - and β - forms. If the OH on the new chiral centre (anomeric carbon) is on the same side of the ring as carbon 6 then the isomer is β and if it is on the opposite side then it is α . A simplification to this is that if the ring is drawn as a Haworth projection with carbon 6 above the ring, then if the OH on the anomeric carbon is below the ring the isomer it is α and if it is above the plane of the ring it is β .

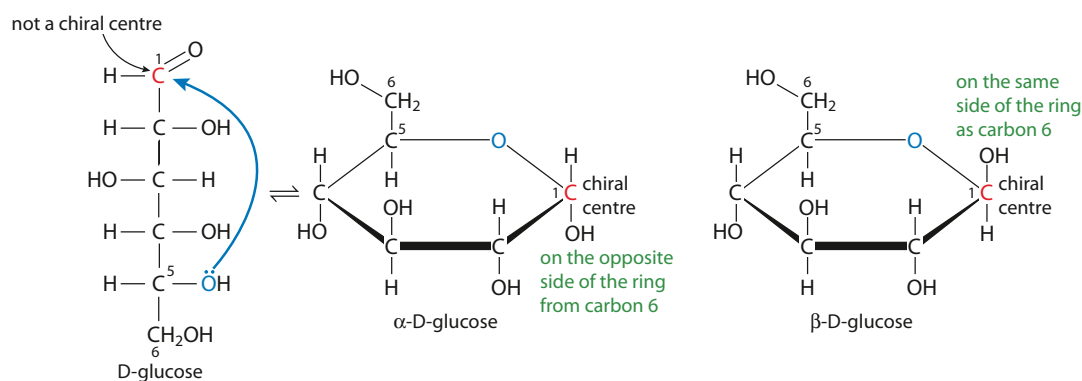


Figure B.99 The formation of α - and β -glucose.

Polysaccharides

Polymers of many monosaccharides joined together are known as polysaccharides. Examples include **starch** (polymer of α -glucose sugars) and **cellulose** (polymer of β -glucose sugars). Note that both these polysaccharides are polymers of glucose.

Cellulose

Cellulose is a polysaccharide responsible for giving structure and strength to plants – it is the most important structural polysaccharide. Cellulose is found in plant cell walls – it makes up about 90% of cotton and about 50% of wood.

It is made up of hundreds of β -glucose units that join together in condensation reactions. The glucose units are linked by β -1,4-glycosidic linkages (Figure B.100). A β -1,4-glycosidic linkage is formed between the C1 (anomeric carbon) of a β -sugar molecule and the C4 of another sugar molecule.

Cellulose is made up of linear chains – it is not branched. The polymer chains can therefore pack closely together. Also, the β -1,4-linkages mean that the most stable conformation of the polymer is a linear one with no coiling – this allows a large number of OH groups to be available for hydrogen bonds with adjacent polymer chains. This means that many hydrogen bonds are formed between the chains, giving the cellulose structure its strength.

The glucose units in cellulose are linked by β -1,4-glycosidic linkages; humans do not possess the enzyme (known as **cellulase**) that hydrolyses these linkages, and therefore we (and most other animals) cannot digest cellulose.

Herbivores such as cows survive on a diet of cellulose-rich plants such as grass. These animals have an extensive digestive system that contains bacteria which secrete cellulase – this allows the breakdown of cellulose into smaller units for digestion.

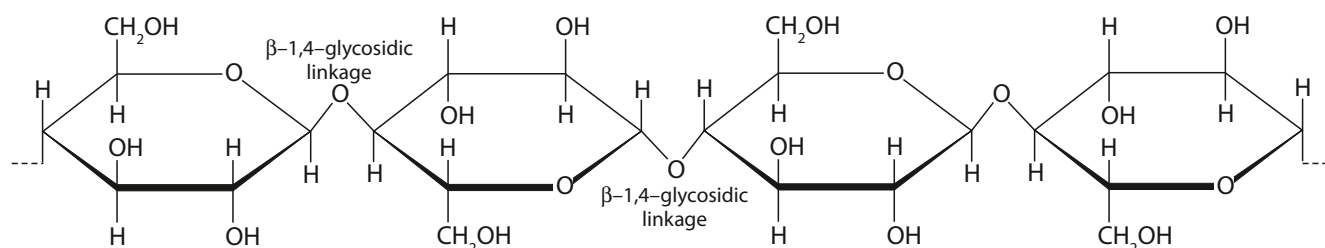


Figure B.100 Glucose units joined together in cellulose.

Cellulose is a major component of dietary fibre. Dietary fibre is plant material that we ingest but are not able to digest. There are two types of dietary fibre:

- **Insoluble fibre** – this includes cellulose, hemicellulose and lignin found in plant cell walls. As it passes through the gut, it binds water and softens and adds bulk to the feces. Dietary sources include whole grains (such as wheat), vegetables and beans.
- **Soluble fibre** – this includes pectin found in plant cells; sources include oats, oatbran and beans.

Dietary fibre has been linked to beneficial effects on a number of conditions/diseases. A diet high in insoluble fibre is useful in treating and preventing constipation, hemorrhoids and diverticulosis (formation of small pouches in the colon) and may improve some cases of irritable bowel syndrome.

A comparison of starch and cellulose

Starch is also a polymer of glucose but of α -glucose units. Starch consists of a mixture of two types of glucose polymers, called **α -amylose** and **amylopectin**. α -amylose consists of thousands of α -glucose units linked together to form linear, unbranched chains. These glucose units are linked by α -1,4-glycosidic linkages (Figure B.101a). An α -1,4-glycosidic linkage is formed between the C1 of an α -sugar molecule and the C4 of another sugar molecule.

Amylopectin is a branched polymer of α -glucose units linked by α -1,4-glycosidic linkages and α -1,6-glycosidic linkages at the branch points (Figure B.101b).

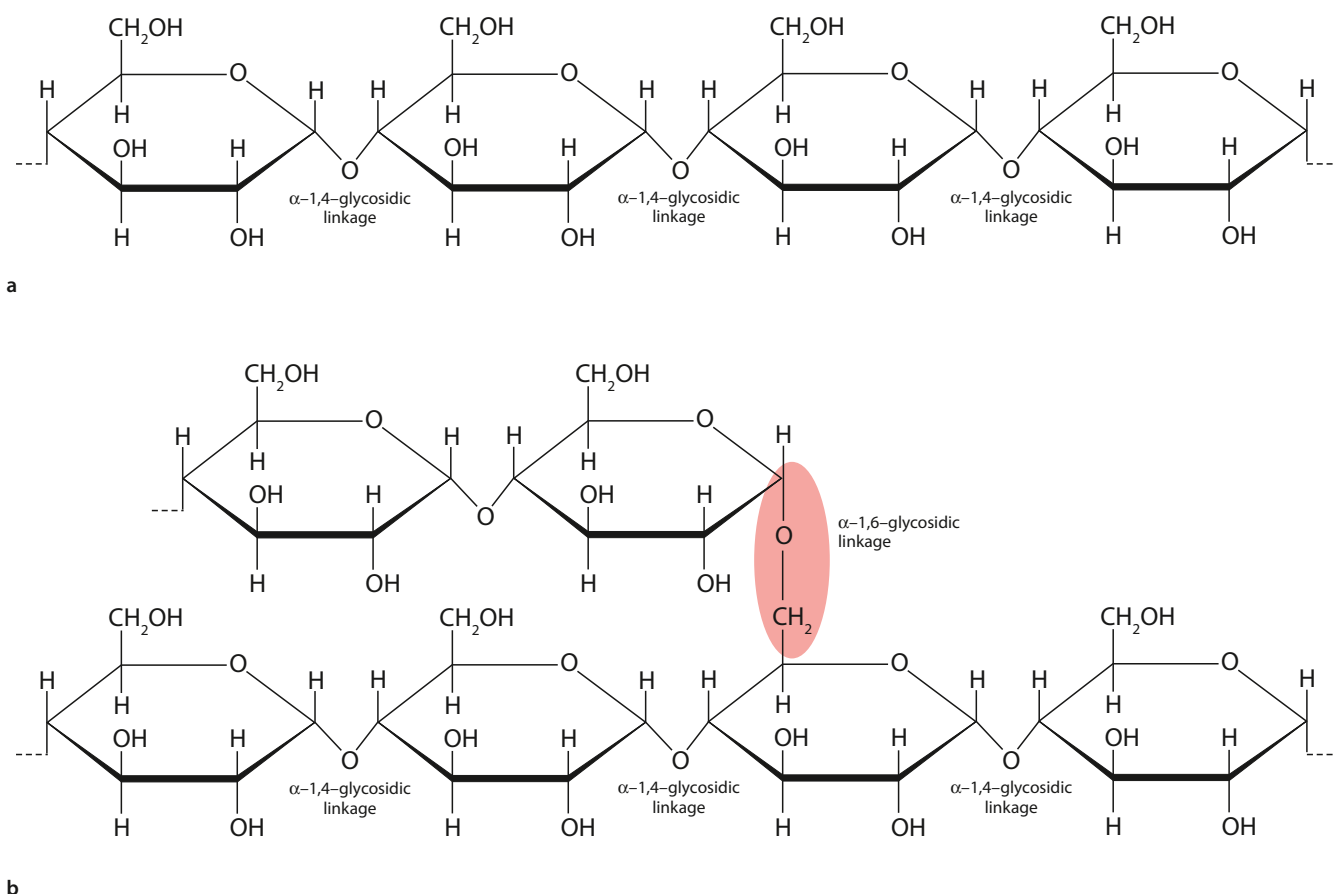


Figure B.101 Examples of how the glucose units are joined in **a** amylose and **b** amylopectin.

The structure of starch granules is quite complex but the presence of α -1,4-glycosidic linkages in the components of starch means that the most stable conformation for the polymer chains is a helical one.

We have digestive enzymes that can **hydrolyse** the α -1,4 and α -1,6 linkages in starch, and hence we can break down the starch polymers into smaller pieces and eventually into single glucose units which can then be absorbed into the bloodstream and used as an energy source.

Table B.4 summarises some of the differences between cellulose and starch.

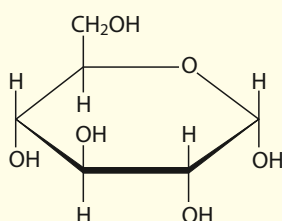
Cellulose	Starch
polymer of β -glucose	polymer of α -glucose
contains β -1,4-glycosidic linkages	contains α -1,4-glycosidic linkages and α -1,6-glycosidic linkages
unbranched chains	amylopectin has branches
linear chains with no coiling	helical chains
strong fibres	not fibrous, not strong
insoluble	soluble in hot water

Table B.4 Some differences between cellulose and starch.

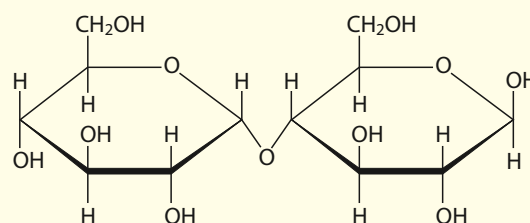
Certain plants have a high starch content in their cells – for example potato tubers, rice grains and wheat. These are the major sources of carbohydrate in the human diet.

? Test yourself

36 Is the following sugar an α -sugar or a β -sugar?



37 A disaccharide is shown below. State the type of reaction that resulted in its formation and name the linkage between the two rings.



Hydrogenated fats

cis-C=C double bonds are much more common in naturally occurring unsaturated fats and oils than *trans*-C=C double bonds.

As discussed earlier (page 22), *cis*-C=C double bonds introduce kinks into the fatty acid chains, which reduces the melting points of fats so that they are liquids at room temperature.

Hydrogenation reactions are used in the food industry to convert C=C double bonds in unsaturated fatty acids to C–C single bonds. In this process, hydrogen gas is added across the double bond(s) present in mono- or, more usually, polyunsaturated vegetable oils at a high temperature (140–225 °C) and pressure in the presence of a nickel, zinc or copper catalyst. This results in what are called **hydrogenated fats** which have higher melting points because of a higher degree of saturation

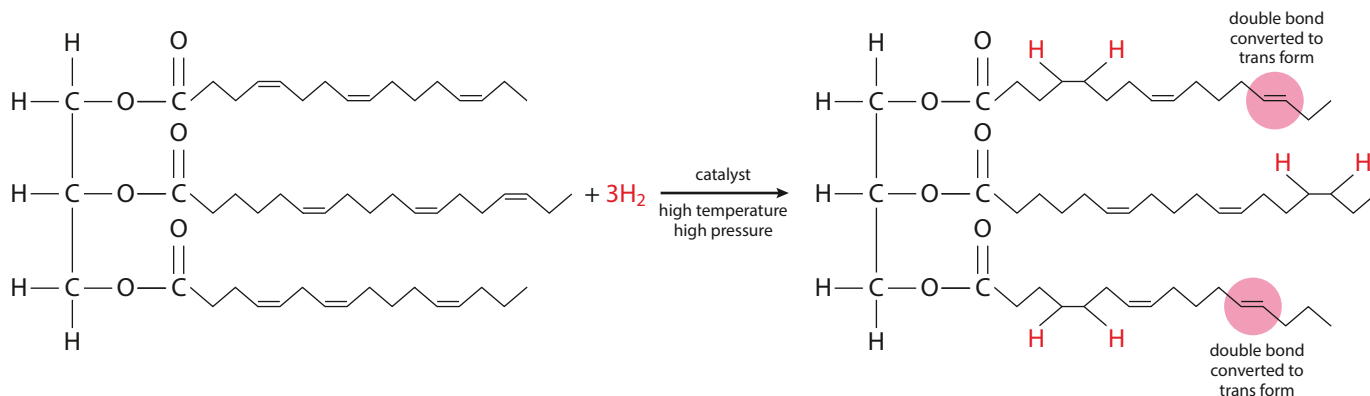


Figure B.102 Partial hydrogenation of a polyunsaturated fat. Only the hydrogen atoms that have been added are shown in the fatty acid chains of the partially hydrogenated product.

(fewer C=C bonds) and are therefore more solid at room temperature. If insufficient hydrogen is added to hydrogenate all the C=C bonds then a partially hydrogenated fat is obtained (Figure B.102).

Hydrogenated fats are used in margarine manufacture (to make oils into solid fats) and in many processed foods.

Advantages of hydrogenation

Partially hydrogenated vegetable oils are cheaper to produce than saturated fats from animals and they also have an increased shelf-life over their fully unsaturated precursors because the rate of oxidation decreases with increased saturation. Hydrogenation also increases hardness and plasticity (stiffness) of the products.

Disadvantages of hydrogenation

Clinical evidence points to health benefits from a diet containing more mono- and polyunsaturated fats than saturated. Saturated fats are more damaging to the heart and circulatory system. In addition, in the hydrogenation process, *cis* fatty acids can isomerise into the *trans* form (Figure B.102). These have been implicated in a number of cardiovascular diseases. The body finds it difficult to metabolise *trans* fatty acids (lipase enzymes seem to recognise only *cis* forms) and so they tend to accumulate in the adipose (fatty) tissue rather than be excreted. *Trans* fatty acids also cause an increase in the more harmful forms of circulating cholesterol – low-density lipoproteins. Finally, *trans* fats are a lower-quality source of energy compared to their *cis* counterparts.



Partially hydrogenated vegetable oils have been banned in some countries.

Vision chemistry

The retina of the eye is made of millions of **rod** cells and **cone** cells. Each of these has a large number of molecules of the protein **rhodopsin** on the surface. Rhodopsin consists of the protein **opsin** bonded to a molecule of *cis*-retinal (Figure B.103a). In *cis*-retinal, one of the C=C double bonds in the chain is *cis*, all the others are *trans*. When a photon of light is absorbed by the retinal in rhodopsin, it changes to the all-*trans* form (Figure B.103b). This changes the conformation of the protein, which triggers a series of events that result in a signal being sent to the brain. Once the

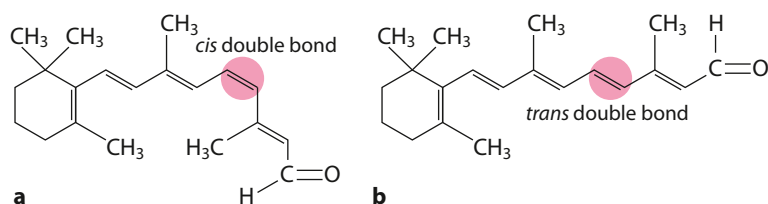


Figure B.103 **a** *cis*-retinal; **b** all-*trans*-retinal.

cis-retinal is also sometimes called '11-*cis*-retinal', where '11' refers to the position of the *cis* double bond.

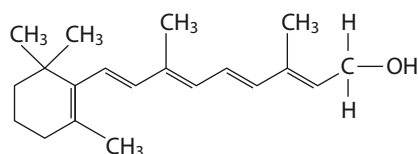


Figure B.104 Vitamin A – retinol.

signal has been sent, the *trans*-retinal dissociates from the opsin and is replaced by another *cis*-retinal molecule. The *trans*-retinal molecule that dissociated is converted to *cis*-retinal again by an enzyme.

Retinal is derived from vitamin A (Figure **B.104**) in a series of enzyme-catalysed reactions. The first reaction involves oxidation of a primary alcohol to an aldehyde and this is followed by an isomerisation reaction to convert one of the *trans*-C=C double bonds to the *cis* form. So, retinal is essential for normal vision and this is why vitamin A deficiency can result in night blindness and, in extreme cases, blindness.

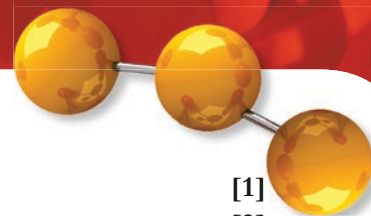
Nature of science

Science can sometimes have unintended consequences. For example, at first sight the development of hydrogenation reactions to convert liquid oils into solid fats that are less prone to going rancid seems like a good advance. As knowledge develops about both the hydrogenation reaction and biochemical processes, unexpected consequences come to light. Nowadays hydrogenated fats are regarded by most people as things to be avoided at all costs.

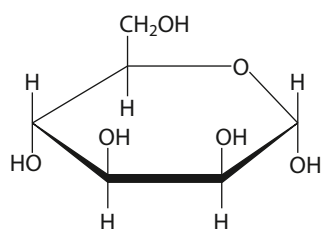
Theories can be used to explain natural phenomena. The understanding that many biological molecules are chiral is essential in understanding the biochemical processes that occur in cells.

Exam-style questions

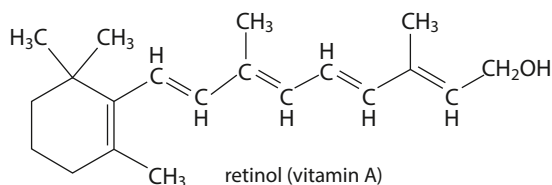
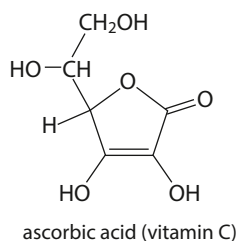
- 1 a Explain the difference between anabolism and catabolism. [2]
- b Photosynthesis and respiration are extremely important reactions in plants.
- i Write overall equations for photosynthesis and respiration, stating clearly which is which. [2]
- ii State whether each of the reactions in part **b i** is a catabolic or anabolic process. [2]
- iii Explain how photosynthesis and respiration are important in maintaining the balance between certain gases in the atmosphere. [2]
- 2 Immunoglobulin G is an antibody that plays an important role in the immune system. It is a protein made up of four polypeptide sub-units.
- a The polypeptides in immunoglobulin G are made up from 2-amino acids.
- i Give the general structural formula for 2-amino acids. [1]
- ii Explain why amino acids have relatively high melting points. [2]
- b Explain what is meant by the primary structure of a polypeptide chain and name the type of bond that links the amino acids together in the primary structure. [2]
- c i Draw the structural formula of **one** of the dipeptides formed when glycine reacts with cysteine (structures are given in Table **B.1**) and name the type of reaction occurring. [2]
- ii Explain why it is possible to form more than one dipeptide when glycine reacts with cysteine. [2]
- d How many possible tripeptides could be produced by reacting together one molecule each of glycine, cysteine and serine? Draw **one** of these tripeptides. [2]
- e The amino acids in immunoglobulin G can be analysed using paper chromatography.
- i Why is the immunoglobulin first treated with dilute hydrochloric acid before paper chromatography is carried out? [1]
- ii Describe how paper chromatography is used to analyse amino acids. [4]
- 3 Lipids play important roles in the body.
- a State **three** roles that lipids play in the body. [3]
- b i Define the term 'iodine number'. [1]
- ii If 11.43 g of I_2 reacts with 0.015 mol of a fatty acid, explain what can be deduced about the structure of this fatty acid. [3]
- c Write an equation to represent the reaction between fatty acids and glycerol to produce a triglyceride (the hydrocarbon chains can be represented by 'R'). [3]
- d Explain how the composition of fatty acids in fats and oils affects their melting point. [3]
- e State the effect of eating high levels of saturated fat on LDL-cholesterol in the body and explain how this can be bad for a person's health. [3]
- 4 Carbohydrates are the most abundant class of biological molecules and include simple sugars as well as complex polysaccharides.
- a State the general formula of carbohydrates and the molecular formula of glucose. [2]
- b Glucose can exist in either straight-chain or ring forms. State the name of a functional group present in the ring form of glucose but not in the straight-chain form. [1]



- c One of the uses of glucose in the body is as an energy reserve.
- i In what form is glucose stored in cells? [1]
 - ii Explain how the structure of the substance named in part c i differs from that of glucose. [2]
- d The structure of a monosaccharide is shown below.



- i Draw the structure of a molecule that is formed when two of these monosaccharide molecules join together. [2]
 - ii Name the class of compound formed when the two monosaccharides join together. [1]
 - iii What is the formula of the other molecule formed in the reaction in part i. [1]
- 5 a The structures of ascorbic acid (vitamin C) and retinol (vitamin A) are shown below. Explain whether they are fat- or water-soluble vitamins. [2]

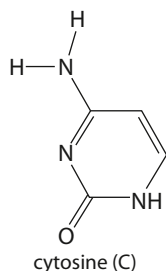
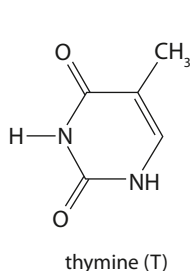
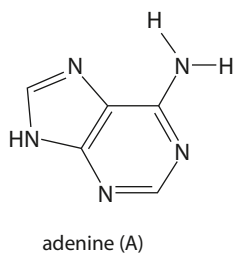
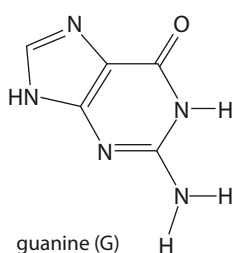


- b Vitamin deficiency is a major problem worldwide. Explain **three** ways of combating vitamin A deficiency in developing countries. [3]
- 6 a Explain what is meant by a xenobiotic and explain **one** environmental problem associated with xenobiotics. [4]
- b State what is meant by the term 'biological detergent' and explain how the use of these is beneficial to the environment. [2]
- c State what is meant by the term 'biomagnification' and explain **one** example of an environmental effect of it. [4]

HL 7 Enzymes catalyse almost every biochemical reaction in the body.

- a** Describe, in general terms, how enzymes catalyse biochemical reactions. [3]
- b** Many medicinal drugs act by inhibiting enzymes in the body, either non-competitively or, more commonly, competitively.
- i** Describe how competitive inhibitors work and how these differ from non-competitive inhibitors. [2]
- ii** What effect does increasing the substrate concentration have on each type of inhibitor? [2]
- iii** How does each type of inhibitor affect V_{\max} and K_m ? [2]
- c** Phosphate buffers are used in biochemical research. A buffer solution was made by dissolving a mixture containing 15.00 g of KH_2PO_4 and 15.00 g of K_2HPO_4 in water and making it up to a total volume of 1.00 dm^3 . Given that the $\text{p}K_a$ for H_2PO_4^- is 7.21, calculate the pH of the buffer solution. [3]

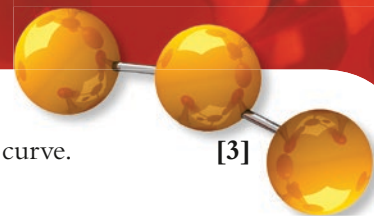
HL 8 a The structures of four nucleotide bases are given below. Draw the **two** sets of base pairs that are found in DNA, including the hydrogen bonds that form between them. [3]



- b** State **two** structural differences between DNA and RNA. [2]
- c** Among some people there is increasing concern about the use of genetically modified food.
- i** Explain what is meant by a genetically modified organism. [2]
- ii** Suggest **two** benefits of genetically modified foods and **two** sources of concern about them. [4]

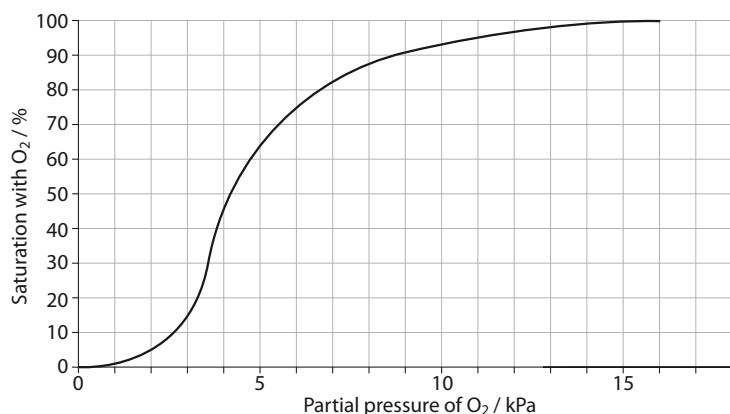
HL 9 Heme is a very important biological pigment and is involved in the transport of oxygen as well as electron-transport reactions. The structure of heme is shown in Figure B.86a and the IB Chemistry data booklet.

- a** State what is meant by a *biological pigment*. [1]
- b** State the oxidation number of iron in deoxygenated hemoglobin. [1]
- c** Explain why heme is coloured, in terms of its structure. [2]



- d The oxygen dissociation curve for hemoglobin is shown below. Explain the shape of this curve.

[3]



- e Explain the effect of increasing the concentration of carbon dioxide and decreasing the pH on the ability of hemoglobin to bind oxygen.

[5]

HL 10 Stereochemistry is very important in biochemical reactions.

- a Draw the structure of the only 2-amino acid that does not have a chiral centre.

[1]

- b State the configuration of all amino acids found in nature.

[1]

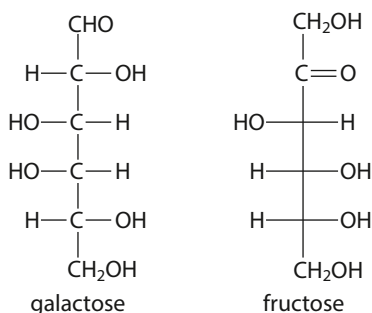
- c Determine whether the following compounds are D or L enantiomers:

i galactose

[1]

ii fructose

[1]



- d Glucose can form two cyclic isomers. Explain why this occurs and name both isomers.

[3]

- e Describe **three** differences between the structures of starch and cellulose.

[3]

- f Describe the role of retinal in vision.

[5]

Option C Energy

C1 Energy sources

C1.1 Energy transformations

Concentrated energy

Energy is the ability to do work. As energy is transferred from one form to another, some does useful work but some will always be lost as heat to the surroundings. The energy that is lost to the surroundings will now be less available to do work and is said to have been **degraded**.

Energy is always transferred in the direction in which it goes from a more concentrated form to a less concentrated (more dispersed) form.

The first law of thermodynamics simply states that energy is conserved – but if energy was transferred without any degradation we would have the basis of a perpetual motion machine because we could continuously transfer energy from one form to another without any losses to a less useful form. In the energy-transfer processes that we will study, some energy will always be lost to the surroundings as low-level heat – this energy is less useful in terms of its potential to do work.

Consider burning coal in a coal-fired power station. Chemical energy is converted to heat energy in a combustion reaction. Some of the heat energy released is dissipated to the surroundings and the rest goes to heat up water to make steam. Some of the kinetic energy of the water molecules in the steam is transferred to a turbine which turns a generator (kinetic energy to electrical energy). In these mechanical components there is friction between moving parts and so some of the energy is again converted to low-level heat energy in the surroundings. The steam then passes through a heat exchanger where it is condensed back to water, with the heat energy being dissipated to the surroundings.

Even the *useful* electrical energy we have obtained from the generator will eventually result in the production of heat – for example, when using some to power a kettle – which will spread out at a low level in the surroundings. When some coal is burned, all the energy released is eventually dispersed in the surroundings – and we cannot get this energy back!

The transfer of heat energy from a hot body to a colder one can be used to do useful work but once the thermal energy is spread out at a low level it is much more difficult to get this energy to do something useful. For instance, there is a lot more energy in a lake at 20 °C than there is in a lump of coal but the energy in the lump of coal is in a much more concentrated form and is therefore more useful. To be useful, the energy has to be in a **concentrated** form.

Learning objectives

- Understand how the quality of energy changes in energy transfer-processes
- Understand what is meant by efficiency and calculate the efficiency of an energy transfer

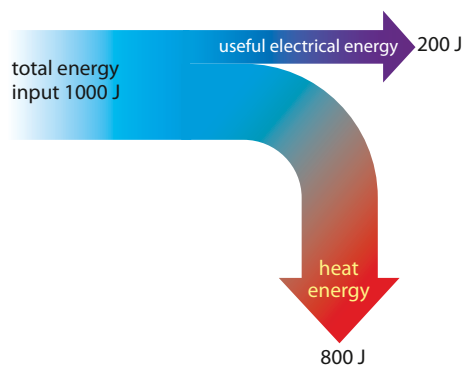


Figure C.1 A Sankey diagram for a solar panel with 20% efficiency.

Efficiency

When comparing energy transformations, it is often useful to know how the amount of useful energy we obtain from a particular transformation compares with the amount of energy put in – the concept of **efficiency** is useful here:

$$\text{efficiency (\%)} = \frac{\text{useful energy out}}{\text{total energy in}} \times 100$$

For instance, if 1000J of light energy from the Sun goes into a solar panel every second and the electrical energy obtained is 200J, then the efficiency of the process is given by:

$$\text{efficiency (\%)} = \frac{200}{1000} \times 100 = 20\%$$

Most of the rest of the energy is transformed into heat and this can be represented in a Sankey diagram (Figure C.1).

? Test yourself

- 1 Calculate the efficiency of each of the following energy transformations:

	Total energy input	Energy output
a		
electric motor	500 J	kinetic energy 350 J heat energy 120 J sound energy 30 J
b		
internal combustion engine	1000 J	kinetic energy 180 J heat energy 750 J sound energy 70 J

- 2 The total kinetic energy of the wind passing through a wind turbine in 1 s is 3000J. The electrical energy obtained per second is 1200J. Calculate the efficiency of the wind turbine.

Learning objectives

- Understand what the desirable characteristics of a useful energy source are
- Understand the difference between renewable and non-renewable energy sources
- Describe the uses of different energy sources
- Discuss the advantages and disadvantages of different energy sources
- Compare different energy sources in terms of energy density and specific energy

C1.2 Energy sources

For a particular energy source to be useful it should have the following characteristics:

- it should release energy at a reasonable rate (not too fast or too slow)
- it should produce minimal pollution.

Other desirable qualities of a fuel include being easy to obtain, and being cheap and plentiful.

Sources of energy can be divided into two categories:

- Renewable energy sources** are naturally replenished – they will not run out.
- Non-renewable energy sources** are finite – they will eventually run out.



Non-renewable energy sources

These are energy sources that will run out in the (near) future and cannot be easily replaced – they are finite. Examples of non-renewable sources of energy are shown in Table C.1.

Source	Description
fossil fuels	Coal, oil and natural gas. These are burned and the heat used to produce steam, which turns a turbine, which turns a generator to produce electricity.
nuclear fission	A nuclear reaction, in which uranium is bombarded with neutrons, produces heat. The heat can be used to generate steam to turn a turbine, which turns a generator to produce electricity.
electrochemical cells	Batteries are portable sources of energy. A redox reaction generates electricity.

Table C.1 Non-renewable forms of energy.

There is a long-running debate about how long supplies of fossil fuels, especially oil, will last. This depends on existing reserves, the discovery of potential reserves, world population, usage and price. There is, however, little doubt about the fact that oil will one day run out and the timescale for this is likely to be measured in tens of years rather than hundreds.

Renewable energy sources

These will not run out because they are replenished naturally. Examples of some renewable energy sources are given in Table C.2.

Source	Description
solar energy	A huge amount of solar energy reaches the Earth every day but only a tiny fraction is used (Figure C.2). Photovoltaic cells convert energy from the Sun to electrical energy.
wind	Wind is used to turn a turbine to generate electricity (Figure C.2).
hydroelectric	Rivers are dammed and water flows from a higher to a lower level through a turbine (change in potential energy) connected to a generator.
tidal	Uses the rising and falling of tides to turn a turbine.
geothermal	Uses heat from rocks underground. Either hot water from underground is used directly or water is pumped down to generate steam, which is used to turn a turbine.
biomass	Energy from plants. This can be used in various ways – such as growing trees and burning them; growing sugar cane and fermenting the sugar to produce ethanol, which can be used as a fuel additive; producing biogas from decaying matter and burning it.
nuclear fusion	Potentially an unlimited supply of energy – joining hydrogen nuclei together with the release of a large amount of energy.

Table C.2 Renewable forms of energy.



Figure C.2 A wind farm and a solar panel site – the solar panels are mounted on motorised columns to track the Sun across the sky.

The term ‘solar panel’ is used in different ways. It is now most often used to describe a panel consisting of linked photovoltaic cells that convert light energy into electrical energy but is also sometimes used to describe a panel that absorbs energy from the Sun to heat water for domestic use.

Advantages and disadvantages of different energy sources

Some advantages and disadvantages of the different energy sources that you will meet in this option are discussed in Table C.3.

Source	Advantages	Disadvantages
solar energy	Energy comes from the Sun – this is free and will not run out in the foreseeable future. Non-polluting in use – e.g. does not produce greenhouse gases. Can be useful in remote locations – e.g. in the desert or on satellites.	Not a very concentrated form of energy and huge banks of solar cells/panels are required. Solar cells and solar panels are expensive to manufacture and buy. Quite a lot of energy and resources needed in manufacture, which can produce pollution. Dependent on the weather and does not produce electricity at night.
burning fossil fuels	Concentrated form of energy – burning fossil fuels can be used to generate the huge amounts of electrical energy required throughout the world each day. Relatively cheap. Gas and oil can be relatively easily transported through pipelines.	Finite – will run out. Greenhouse gases are generated, which can contribute to climate change. Other pollutants such as NO_x and SO_2 can be produced; these lead to acid deposition. Potential for environmental disasters through extraction of crude oil (e.g. Gulf of Mexico, 2010) and transportation by ship (e.g. Sea Empress, 1996). Coal mining – opencast mining can destroy large areas of land and can be dirty and unsightly for local residents; underground mining can be dangerous.
nuclear fission	Very concentrated form of energy. Does not produce greenhouse gases in operation (although some are produced when fuel is mined/transported to the power station).	Nuclear waste is produced, which is difficult to dispose of safely. Danger of an accident – e.g. Chernobyl, 1986, and Fukushima, 2011 – and the release of radioactivity into the environment. Difficult and expensive to decommission nuclear power stations.
non-rechargeable batteries	They are portable.	They are expensive. Have only a limited lifetime – they run out. Problems with disposal – many contain heavy metals.
rechargeable batteries	They are portable. Can be used over and over again.	They are expensive to buy. They run out and have to be recharged every so often.
hydrogen fuel cells	They are portable. They do not use fossil fuels. They use hydrogen, which can be extracted from water, which is present in virtually limitless supply. They do not produce carbon dioxide in operation.	There are difficulties with storing hydrogen (a flammable gas). How environmentally friendly they are depends on how the hydrogen is produced.
nuclear fusion	Has the potential to produce vast amounts of energy. Does not produce greenhouse gases. Does not produce nuclear waste.	There is no working nuclear fusion power plant – it is very difficult to produce and contain a sustained nuclear fusion reaction that generates more energy than needs to be put in to produce the extremely high temperatures required.

Table C.3 Advantages and disadvantages of different energy sources.



Comparing of the amounts of energy released in combustion reactions

When fuels are burned, energy is released in the form of heat (chemical energy converted to thermal energy). **Specific energy** and **energy density** are sometimes used to compare different fuels.

- $\text{Specific energy} = \frac{\text{energy released from fuel}}{\text{mass of fuel consumed}}$
- $\text{Energy density} = \frac{\text{energy released from fuel}}{\text{volume of fuel consumed}}$

We can calculate the specific energy of hydrogen from its enthalpy change of combustion. The enthalpy change of combustion is the enthalpy change when one mole of substance undergoes complete combustion. For hydrogen the value is -286 kJ mol^{-1} .

One mole of hydrogen (H_2) has a mass of 2.02 g, so burning 2.02 g of hydrogen produces 286 kJ of energy.

$$\text{specific energy} = \frac{286}{2.02} = 142 \text{ kJ g}^{-1}$$

This could also be expressed in other units, such as MJ kg^{-1} . Because there are 1000 kJ in 1 MJ and 1000 g in 1 kg, 142 kJ g^{-1} is equivalent to 142 MJ kg^{-1} .

To calculate hydrogen's energy density, we need to know the volume occupied by one mole of hydrogen.

At 25°C and 100 kPa, the volume occupied by one mole of an ideal gas is 24.8 dm^3 (this can be calculated using $PV = nRT$) and so the energy density of hydrogen can be calculated:

$$\text{energy density} = \frac{286}{24.8} = 11.5 \text{ kJ dm}^{-3}$$

This is equivalent to 11.5 J cm^{-3} or 11.5 MJ m^{-3} .

We can compare these values with those for octane (C_8H_{18}) – a liquid fuel that is a constituent of gasoline:

$$\text{specific energy} = \frac{5470}{114.26} = 47.87 \text{ kJ g}^{-1} \text{ or } 47.87 \text{ MJ kg}^{-1}$$

Octane has a density at 25°C of 0.703 g cm^{-3} , and because

$\text{density} = \frac{\text{mass}}{\text{volume}}$ we can calculate the volume occupied by one mole of octane:

$$\begin{aligned} \text{volume} &= \frac{114.26}{0.703} \\ &= 163 \text{ cm}^3 \text{ or } 0.163 \text{ dm}^3 \end{aligned}$$

$$\begin{aligned} \dots \text{ and the energy density} &= \frac{5470}{0.163} \\ &= 33\,600 \text{ kJ dm}^{-3} \end{aligned}$$

Energy density is sometimes used to describe the energy released per unit mass, rather than per unit volume – it is important to check the units to make sure what the term is describing.

1 MJ is $1 \times 10^6 \text{ J}$.

We usually use the molar volume measured at STP ($22.7 \text{ dm}^3 \text{ mol}^{-1}$) but the problem here is that the standard enthalpy change of combustion is measured at 298 K. We could have made the assumption that the enthalpy change of combustion does not vary much with temperature, in which case we could have calculated the energy density as $\frac{286}{22.7} = 12.6 \text{ kJ dm}^{-3}$

114.26 g is the mass of one mole of octane.

A slightly different answer is obtained if more significant figures are carried through on the calculator.

It can be seen from these values that when hydrogen burns it releases approximately three times more energy per gram than octane, but the energy released per unit volume is more than 2000 times higher for octane. This illustrates a problem with using hydrogen as a fuel – the storage of a large volume of a highly flammable gas.

A typical family car could have an average fuel consumption of 7 litres (dm^3) of gasoline per 100 km – so to travel the same distance more than 1500 litres (dm^3) of hydrogen gas would be required (the size of a trunk/boot of a car is typically about 400 dm^3). One solution to this problem would be to store the hydrogen under pressure but this requires fuel tanks to be made of thicker material which is also heavier and more expensive. Other methods of storage are being investigated such as cryo-compression (storing at low temperature under pressure) and various forms of chemical storage.

When considering the best fuel for a particular job, both the specific energy and the energy density must be considered. When weight is a problem then specific energy is likely to be more important – for instance, the space shuttle used liquid hydrogen as a fuel. When storage volume for the fuel is an issue then energy density becomes more important.

Nature of science

All science has to be funded – either by governments, international organisations or companies. Science should, however, not be biased or be influenced by the vested interests of large international corporations such as oil companies. Scientists can sometimes be put in difficult positions when their findings do not fit with the interests of the people funding their research.

We can understand the non-existence of perpetual motion machines by describing energy as having both quantity and quality. In an energy transfer the total amount of energy transferred is constant but the quality is degraded.

? Test yourself

- 3 Calculate the specific energy of each of the following fuels:

	Fuel	Formula	Enthalpy change of combustion / kJ mol^{-1}
a	methanol	CH_3OH	–726
b	hexane	C_6H_{14}	–4163
c	benzaldehyde	$\text{C}_6\text{H}_5\text{CHO}$	–3525

- 4 Calculate the energy density of each of the following fuels – give all your answers in kJ dm^{-3} :

	Fuel	Formula	Physical state at 298 K and 100 kPa	Enthalpy change of combustion / kJ mol^{-1}	Molar volume (dm^3) at 100 kPa and 298 K	Density / g cm^{-3}
a	methane	CH_4	gas	–891	24.8	–
b	ethyne	C_2H_2	gas	–1301	24.8	–
c	benzene	C_6H_6	liquid	–3525	–	0.877
d	ethanol	$\text{C}_2\text{H}_5\text{OH}$	liquid	–1367	–	0.789

C2 Fossil fuels

C2.1 Formation of fossil fuels

The main fossil fuels

The three main fossil fuels are **coal**, **oil** and **natural gas**. They are called **fossil fuels** because they are formed from things that were once alive and have been buried underground for millions of years.

Coal is formed from the remains of plants and trees which fell into swamps millions of years ago. These plants were then covered in layers of sediment and underwent partial decomposition in the absence of oxygen, and under high pressure and temperature, to form coal. Coal can contain up to 95% carbon by mass. The best coals are those with the highest percentage of carbon – these burn most cleanly and release the greatest amount of heat on combustion. Other elements that may be present in coal are hydrogen, oxygen, nitrogen and sulfur.

Crude oil (petroleum) was formed from marine organisms (plankton) that died millions of years ago and sank to the bottom of the sea. These were covered by layers of sediments and underwent chemical processes under conditions of high pressure, moderate heat (between about 60 °C and 170 °C) and the absence of oxygen to convert them into crude oil. Crude oil is a complex mixture of hydrocarbons, including straight and branched-chain alkanes, cycloalkanes and aromatic compounds. Other elements that may be present are nitrogen, oxygen and sulfur.

Natural gas is formed in basically the same way as crude oil and is often found with it. Natural gas consists mostly of methane. Other light hydrocarbons may also be present, as well as hydrogen sulfide.

Fossil fuels are formed by the reduction of biological compounds.

That the reactions involved in the formation of fossil fuels are reduction reactions can be seen by looking at the differences in the oxidation numbers of carbon in biological compounds and fossil fuels. Biological compounds in living organisms that will eventually be converted to fossil fuels include proteins (polymers formed from amino acids) and carbohydrates.

The simplest amino acid is 2-aminoethanoic acid (glycine), the structure of which is shown in Figure C.3. If it is assumed that oxidation numbers are oxygen -2 , hydrogen $+1$ and nitrogen -3 , the average oxidation number of carbon in 2-aminoethanoic acid is $+1$. The oxidation number of carbon in methane is -4 and in coal (assuming that it is pure carbon) is 0 . So, conversion of the amino acid into a fossil fuel involves reduction.

Similarly, the average oxidation number of carbon in glucose ($C_6H_{12}O_6$) is 0 and that in octane (C_8H_{18}) is -2.25 .

Learning objectives

- Understand what is meant by a fossil fuel
- Understand that fossil fuels were formed from biological compounds that contained carbon, hydrogen, nitrogen, sulfur and oxygen
- Understand that fossil fuels were formed in reduction reactions
- Discuss the advantages and disadvantages of fossil fuels

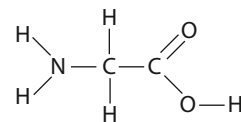


Figure C.3 The simplest amino acid – 2-aminoethanoic acid.

It can also be seen that oxygen is removed when glucose is converted into octane – removal of oxygen is reduction.

It is very difficult to estimate reserves of coal, oil and gas and how long they will last. Factors such as undiscovered reserves, how usage will change, new technologies etc. all have to be considered.

Advantages and disadvantages of fossil fuels

Table C.4 gives details of some advantages and disadvantages of the three main fossil fuels.

Of these fuels, natural gas is currently the cheapest to produce and there is a trend in the USA away from coal-fired power stations to natural-gas electricity generation, driven by price.

Fuel	Advantages	Disadvantages
coal	Supplies should last hundreds of years. Distributed throughout the world. Can be converted into synthetic gaseous and liquid fuels. Can be converted into feedstock for the petrochemical industry. More concentrated source of energy than most renewable sources of energy. Relatively cheap to produce.	Produces greenhouse gases when burned so contributes to climate change. Not as easily transported as oil or gas. Mining is a dirty (slag heaps) and dangerous process. Dirty fuel – pollutants include particulates (fly ash) and sulfur dioxide (acid rain).
oil	Easily transported in pipelines or by tankers. Convenient fuel for internal combustion engine. Source of variety of chemicals for the petrochemical industry.	Produces greenhouse gases when burned so contributes to climate change. Very limited lifespan – supplies could run out in decades. Can produce sulfur dioxide (acid rain) when burned. Environmental problems associated with extraction and transportation in tankers. Only a few countries have reserves.
gas	It is a clean fuel. Easily transported in pipelines and tankers. Releases a higher quantity of energy per kg than coal or oil. Produces less CO ₂ per kJ of energy released. Cheap to produce – comes as a byproduct of coal and oil production.	Produces greenhouse gases when burned so contributes to climate change. Limited lifespan – may be less than 100 years. Only certain countries have reserves. Risk of explosions due to leaks. More difficult to store than coal and oil because it is a gas – must be stored under pressure or cooled to liquefy it.

Table C.4 Advantages and disadvantages of fossil fuels.



In the 1970s people were saying that crude oil would run out in the next 40 years but now, 40 years on, this is still being said. Are these numbers just being used for effect, are estimates being revised in the light of new data or were the scientists in the 1970s simply wrong? There are many different estimates of how long the supplies of crude oil will last; how do we know what to believe?

Learning objectives

- Understand that petroleum is a mixture of hydrocarbons
- Understand that petroleum must be split into fractions before use
- Recall the names of fractions, their relative volatility and uses

C2.2 Petroleum refining

Fractional distillation of crude oil

Petroleum (crude oil) is a complex mixture of thousands of different hydrocarbons.

These hydrocarbons may be alkanes, cycloalkanes (ring compounds) or aromatic compounds (containing benzene rings).

Petroleum is not composed solely of hydrocarbons – it also contains a small proportion of compounds containing nitrogen, oxygen and/or sulfur, as well as carbon and hydrogen.

Petroleum can be separated into a series of simpler mixtures called **fractions** by the process of fractional distillation. The separation process relies on the different components of the mixture having different boiling points so that they condense at different levels in the fractionating column. This process is carried out in an oil refinery and is called **refining**.

The crude oil is heated to about 350 °C in a furnace (Figure C.4). This is high enough to vaporise most components of the petroleum. The liquid–vapour mixture is passed into a fractionating tower which is hot at the bottom and cooler at the top. The lowest-boiling components do not condense at all in the tower and are drawn off at the top as gases (refinery gases). Compounds of intermediate boiling point travel up the column until they condense and are drawn off as liquids. Lower-boiling-point fractions are drawn off higher up the tower. The highest-boiling-point components, which were not vaporised in the furnace, sink to the bottom of the tower and are drawn off there. This fraction, normally just called the **residue**, can be further separated into other fractions using vacuum distillation.

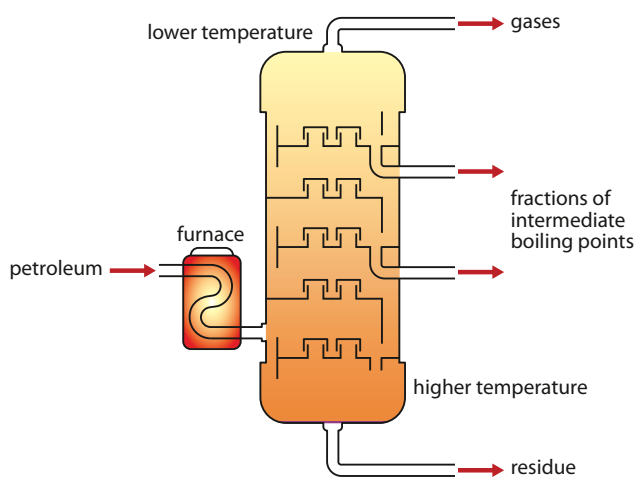


Figure C.4 Crude oil can be separated into fractions using fractional distillation.

Fractional distillation involves separation by physical processes (boiling and condensing) and relies on the different physical properties (boiling points) of the components.

All the fractions obtained from petroleum contain a mixture of compounds and boil over a range of temperatures. For instance, gasoline typically contains hydrocarbons with between five and nine carbon atoms with boiling points in the range 40–200 °C, whereas kerosene contains compounds which have, on average, more carbon atoms per molecule and boils between about 150 °C and 300 °C. Table C.5 shows the names and uses of some fractions.

As the number of carbon atoms in a molecule increases, the strength of the London forces between the molecules increases and the compounds become less volatile (evaporate less easily).

Exam tip

'Crude oil' and 'petroleum' are used interchangeably in the syllabus.

Petroleum has virtually no uses until it is separated into fractions.

Fraction	Uses			
refinery gases	Fuel for cooking and heating. Bottled gas. Used for fuel in the refinery.			
gasoline	Fuel for cars.			
naphtha	Feedstock for petrochemical industry. Converted by catalytic reforming into gasoline. Used as a solvent.			
kerosene (paraffin)	Jet fuel. Household heaters and lamps. May also be cracked to produce more gasoline.			
diesel oil (gas oil)	Diesel fuel for cars, lorries etc. May be cracked to produce more gasoline.			
fuel oil	Fuel for ships and industry. Fuel for home central heating systems.			
lubricating oil	Lubricant in engines and machinery. May be cracked.			
wax	Candles, petroleum jelly, waxed paper and cardboard in the food industry.			
bitumen/asphalt	Tarmac for roads and waterproofing roofs.			

Table C.5 Fractions obtained from crude oil. Other names are sometimes used for fractions. Volatility refers to how readily a substance evaporates.

Learning objectives

- Understand that an octane number is a measure of the tendency of a fuel not to auto-ignite
- Understand how octane numbers are related to molecular structure
- Understand that cracking and reforming may be carried out to improve the octane number of a hydrocarbon fraction
- Write equations for cracking and reforming reactions

Octane number is sometimes called 'octane rating'.

C2.3 Octane numbers

The auto-ignition problem

In an internal combustion engine, a fuel–air mixture is compressed in a cylinder and ignited by a spark from a spark plug. The flame spreads smoothly through the cylinder and only a small fraction of the fuel is burning at any one time. However, depending on the fuel used, it can also happen that the fuel explodes in the cylinder – it all burns at essentially the same time. This mini-explosion, often called 'auto-ignition', results in a metallic knocking sound – called 'knocking' or 'engine knock'. Knocking leads to wear in the engine and wastage of petrol.

The tendency of a fuel to auto-ignite and cause knocking depends on the molecular structure of the components of the petrol used. Gasoline mixtures that are rich in straight-chain alkanes, such as heptane, have a high tendency to auto-ignite and cause knocking. The combustion of branched-chain alkanes, such as 2,2,4-trimethylpentane, is much smoother and more controlled. Gasoline mixtures rich in branched-chain alkanes are more efficient fuels and are less likely to cause knocking.

It is convenient to have a measure of the suitability of alkanes as fuels and this is why each compound is given an **octane number**, determined by experiment. Two arbitrary reference points are used in the scale – heptane, which is assigned an octane number of 0; and 2,2,4-trimethylpentane, which is assigned an octane number of 100 (Figure C.5).

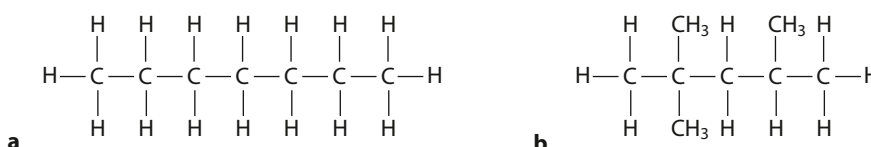


Figure C.5 **a** Heptane, octane number 0, is a poor fuel – it has a high tendency to knock; **b** 2,2,4-trimethylpentane, octane number 100, is a good fuel – it has a low tendency to knock.



The test fuel is placed in a test engine and the compression ratio increased until knocking occurs. The test fuel is then replaced by a series of reference fuels – mixtures of heptane and 2,2,4-trimethylpentane – until a fuel is found that has the same tendency to auto-ignite/knock as the test fuel. So an octane number refers to the performance of a fuel relative to heptane and 2,2,4-trimethylpentane.

An octane number is a measure of the tendency of a fuel not to undergo auto-ignition (cause knocking) in an engine. The higher the octane number, the lower the tendency to undergo auto-ignition/cause knocking.

Modern car engines usually need petrol with an octane number of 90+.

The octane number is affected by various structural features:

- straight-chain alkanes have the lowest octane number and this number decreases as the number of carbon atoms in the chain increases – for instance, pentane has an octane number of 62 and hexane an octane number of 25
- branched-chain alkanes have higher octane numbers than straight-chain alkanes – for example, 2-methylhexane, an isomer of heptane (octane number 0), has an octane number of 44
- as the degree of branching increases so does the octane number – 2,2-dimethylpentane has an octane number of 93, whereas its isomer with three branches, 2,2,3-trimethylbutane, has an octane number of 113
- straight-chain alkenes have higher octane numbers than straight-chain alkanes
- cycloalkanes have higher octane numbers than straight-chain alkanes
- aromatic compounds generally have the highest octane numbers.

Increasing the octane number of fuels

The original solution to increasing the octane number of fuels was to add an anti-knock compound, the most common of which was lead tetraethyl [$\text{Pb}(\text{C}_2\text{H}_5)_4$]. However, concerns about the toxicity of lead have resulted in leaded gasoline being phased out in most countries. In addition, leaded fuels cannot be used with catalytic converters because they poison the catalyst. Other solutions have included adding MTBE (methyl tertiary-butyl ether) but there has been growing concern about its use in the US because of contamination of drinking water – contamination with MTBE makes the water undrinkable (unpleasant taste and odour). It is also suspected that MTBE could be a human carcinogen. MTBE has been largely replaced by ethanol in the US but not in other countries. Another solution to improving the octane number of gasoline is to increase the concentration of benzene and other aromatic compounds. However, benzene is also a known human carcinogen and therefore most countries set limits on the concentration of benzene in fuel. Not all countries, however, set limits and these limits vary considerably between countries.

Catalytic cracking

Catalytic cracking increases the yield of gasoline from a refinery by breaking down long-chain molecules from the less valuable fractions obtained from petroleum into shorter ones. Shorter-chain alkanes

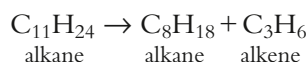
‘Compression ratio’ is the ratio of the volume of an engine cylinder when the piston is at the bottom of its stroke compared to when it is at the top of its stroke – the higher the compression ratio, the greater the likelihood of knocking.



The numbers used here are all research octane numbers (RON) but other octane numbers are also used, such as the motor octane number (MON), blending octane number (BON), pump octane number (PON) and anti-knock index (AKI). Different numbers are used in different countries.

have a higher octane number than longer-chain ones and so cracking increases the octane number of a petroleum fraction. Catalytic cracking also increases the octane number of gasoline because it produces more branched-chain alkanes and aromatic compounds.

The hydrocarbon molecules to be cracked are passed through a bed of a zeolite (aluminosilicate) catalyst at about 500 °C under pressure. The reactions involved in catalytic cracking are complex but a representative reaction is:



At the simplest level, cracking will produce an alkane and an alkene because there are not enough hydrogen atoms in the original molecule to produce two alkanes. Various isomerisation reactions can occur during this process to produce branched-chain alkanes, cycloalkanes and aromatic compounds which increase the octane number.

A variation on the above process is **hydrocracking** which involves heating the long-chain hydrocarbons with hydrogen and a catalyst at high temperature and pressure. This converts the long-chain alkanes into shorter-chain alkanes but no alkenes, due to the presence of the hydrogen. Branched alkanes and cycloalkanes are also produced, which increase the octane number of the fuel.

Thermal cracking

Long-chain molecules may also be cracked simply by heating them. This process is known as **thermal cracking** or **steam cracking**. Typically this uses a high temperature (about 900 °C) and sometimes also high pressure to break down long-chain alkanes into shorter-chain alkanes and alkenes. The alkenes have a higher octane number than alkanes but are not desirable in car engines because they can form gums which can clog up the engine – but they are very useful as a chemical feedstock for making polymers, for example.

Catalytic reforming

Catalytic reforming is a chemical process that increases the octane number of a petroleum fraction by increasing the proportion of branched-chain alkanes, cycloalkanes and aromatic compounds. This is achieved by heating the hydrocarbon feedstock to about 500 °C in the presence of a catalyst such as platinum on an alumina support. Many different reactions can occur – some examples are given in Table C.6.



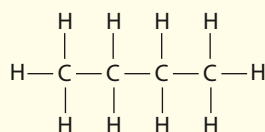
isomerisation	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3 \rightarrow \begin{array}{ccccccc} & \text{H} & & \text{CH}_3 & & \text{H} & & \text{H} & & \text{H} \\ & & & & & & & & & \\ \text{H} & -\text{C} & - & \text{C} & - & \text{C} & - & \text{C} & - & \text{C} & - & \text{H} \\ & & & & & & & & & \\ & \text{H} & & \text{H} & & \text{H} & & \text{H} & & \text{H} \end{array}$ <p style="text-align: center;">hexane 2-methylpentane</p>
cyclisation	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3 \rightarrow \begin{array}{ccccc} & \text{H} & & \text{H} & \\ & & & & \\ \text{H} & -\text{C} & & \text{C} & -\text{H} \\ & & & & \\ \text{H} & -\text{C} & & \text{C} & -\text{H} \\ & & & & \\ & \text{H} & & \text{H} & \end{array} + \text{H}_2$ <p style="text-align: center;">hexane cyclohexane</p>
aromatisation	$\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3 \rightarrow \text{C}_6\text{H}_6 + 4\text{H}_2$ <p style="text-align: center;">hexane benzene</p>

Table C.6 Some reactions that can occur in catalytic reforming.

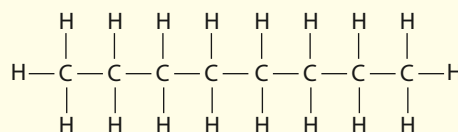
? Test yourself

5 From each pair select the molecule with the higher octane number:

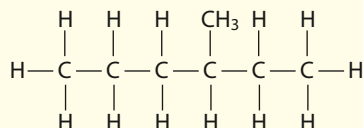
a M



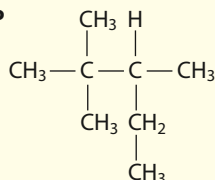
N



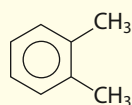
b O



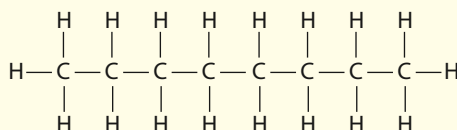
P



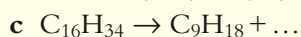
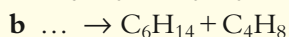
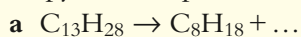
c Q



R



6 Copy and complete the following equations for cracking reactions:



Learning objectives

- Understand what is meant by coal gasification and liquefaction
- Write equations for the processes involved in coal gasification and liquefaction

Exam tip

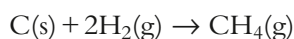
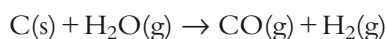
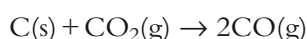
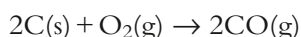
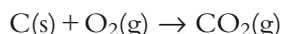
If asked to write just one equation for the coal gasification reaction

$\text{C(s)} + \text{H}_2\text{O(g)} \rightarrow \text{CO(g)} + \text{H}_2\text{(g)}$ is probably the most appropriate.

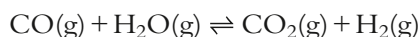
C2.4 Coal gasification and liquefaction

Coal gasification

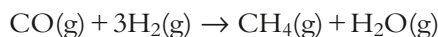
This refers to the process of converting coal (mostly carbon) into a gaseous fuel. The gasification process usually involves conversion of coal into a mixture of carbon monoxide and hydrogen – a mixture known as **syngas** (synthesis gas). There may also be small amounts of methane present. Coal is heated to high temperatures (above 1000 °C) with oxygen and steam; high pressures may also be used. Various reactions occur in the coal gasification process:



Syngas of various compositions (CO : H₂ ratio) is needed for different processes and the **water–gas shift reaction** is used to control this. If syngas is treated with steam, some carbon monoxide reacts with the steam to reduce the carbon monoxide composition relative to H₂:



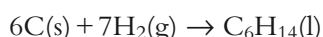
Syngas (which has been purified by removal of sulfur compounds) can be converted into synthetic natural gas (substitute natural gas, SNG) by heating it with more hydrogen:



Coal is more plentiful than natural gas and this reaction can be used to increase supplies of natural gas. The product is a clean-burning fuel (impurities such as sulfur are removed) which is easier to transport.

Coal liquefaction

Coal liquefaction refers to the process of converting coal to liquid hydrocarbons. Coal may be liquefied either directly or indirectly. In the direct processes the coal is mixed with a solvent and hydrogen and subjected to high temperature (around 400 °C) and high pressure in the presence of a catalyst (such as ‘red mud’, which is a byproduct of the processing of bauxite for the extraction of aluminium). This can result in reactions such as:



C₆H₁₄ is a liquid at room temperature and pressure.

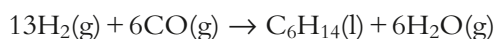
The indirect process involves the conversion of syngas (mostly CO and H₂) from the gasification process into liquid hydrocarbons using the Fischer–Tropsch process. Different plants use different conditions but a typical set of conditions would be a temperature of about 300 °C, high



The use of chemical technology can be driven by political factors. Germany used the Fischer–Tropsch process during World War II to produce liquid hydrocarbons to power vehicles; South Africa used the same technology 30 years later when it was largely isolated by trade sanctions during the apartheid era.



pressure and a catalyst (e.g. ruthenium on an alumina support). Alkanes can be formed by reactions such as:



This can be generalised as $(2n+1)\text{H}_2 + n\text{CO} \rightarrow \text{C}_n\text{H}_{2n+2} + n\text{H}_2\text{O}$.

C2.5 Carbon footprints

Carbon dioxide production from burning fossil fuels

All fossil fuels contain carbon and when they are burned the carbon is converted into carbon dioxide (and carbon monoxide and soot). Carbon dioxide produced by burning fossil fuels is believed to contribute to climate change and so it is important to be able to compare the amounts of carbon dioxide produced when different fuels are burned.

We can use the enthalpy change of combustion values given in Table C.7 to compare three fossil fuels in terms of the amount of carbon dioxide per gram of fuel burned and per kJ of energy released.

Substance	Enthalpy change of combustion / kJ mol^{-1}
coal – C(s)	–394
natural gas – $\text{CH}_4(\text{g})$	–891
petrol – $\text{C}_8\text{H}_{18}(\text{l})$	–5470

Table C.7 The enthalpy change of combustion of fossil fuels.

The equations for the complete combustion of each of these fuels are:

- $\text{C}(\text{s}) + \text{O}_2(\text{g}) \rightarrow \text{CO}_2(\text{g})$
- $\text{CH}_4(\text{g}) + 2\text{O}_2(\text{g}) \rightarrow \text{CO}_2(\text{g}) + 2\text{H}_2\text{O}(\text{l})$
- $\text{C}_8\text{H}_{18}(\text{l}) + \frac{25}{2}\text{O}_2(\text{g}) \rightarrow 8\text{CO}_2(\text{g}) + 9\text{H}_2\text{O}(\text{l})$

We can now calculate the amounts of carbon dioxide produced when one mole of each fuel is burned:

- 1 mol carbon produces 1 mol (44.01 g) CO_2
- 1 mol methane produces 1 mol (44.01 g) CO_2
- 1 mol octane produces 8 mol (352.08 g) CO_2 .

If we divide each value by the molar mass of the fuel burned, we get the mass of CO_2 produced per gram of fuel burned:

- carbon $\frac{44.01}{12.01} = 3.66$ g per gram of fuel burned
- methane $\frac{44.01}{16.05} = 2.74$ g per gram of fuel burned
- octane $\frac{352.08}{114.26} = 3.08$ g per gram of fuel burned

Learning objectives

- Understand what is meant by a carbon footprint
- Calculate how much carbon dioxide is added to the atmosphere when different fuels are burned
- Calculate carbon footprints

We are using a simplified model here to make the calculations easier. We are assuming that coal is made entirely of carbon, that natural gas is pure methane and that petrol is pure octane.

When we are considering energy production, perhaps a more useful comparison is the mass of CO₂ produced per kJ of energy released. This value can be obtained by dividing the mass of CO₂ produced per mole by the enthalpy change of combustion:

- carbon $\frac{44.01}{394} = 0.112 \text{ g of CO}_2 \text{ per kJ of energy released}$
- methane $\frac{44.01}{891} = 0.0494 \text{ g of CO}_2 \text{ per kJ of energy released}$
- octane $\frac{352.08}{5470} = 0.0644 \text{ g of CO}_2 \text{ per kJ of energy released}$

Because coal is composed entirely of carbon it comes out worst in all these calculations.

Carbon footprint

A **carbon footprint** is a measure of the total amount of greenhouse gases (primarily carbon dioxide and methane) emitted as a result of human activities. It is usually expressed as equivalent tonnes of carbon dioxide (e.g. 10 tonne CO₂e). Carbon footprints can be worked out at many different levels – for example by country, region, organisation, household or individual. Your carbon footprint is influenced by many things – major contributions come from transport and electricity usage but what food you eat, whether you buy a newspaper etc. all make a difference.

Contributions to an individual's carbon footprint can be divided into two broad categories – direct production and indirect production. Direct production of carbon dioxide from, say, a car journey is reasonably straightforward to calculate but indirect production from something like eating a frozen pizza bought from a supermarket is much more difficult to quantify.

Worked example

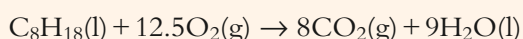
C.1 Work out the carbon footprint for a car journey of 100 km. Assume that the car uses 7 dm³ of fuel for the journey and that the fuel is octane (C₈H₁₈).

The mass of octane burned can be calculated from its density (0.703 g cm⁻³):

$$\begin{aligned}\text{mass} &= \text{density} \times \text{volume} \\ &= 0.703 \times 7000 \\ &= 4921 \text{ g}\end{aligned}$$

The number of moles of octane is $\frac{4921}{114.26} = 43.07 \text{ mol}$

The equation for the combustion of octane is:



So $8 \times 43.07 = 345 \text{ mol CO}_2$ is produced.

The mass of CO₂ produced is $345 \times 44.01 = 15\,200 \text{ g}$ or 15.2 kg



Working out a carbon footprint from using electricity is more complicated because it depends on how the electricity is generated. The carbon footprint from watching television for one hour a day for a year can be estimated by making some assumptions.

The power rating of a typical modern TV set is about 100W, so the approximate annual energy use is $0.10 \text{ kW} \times 365 \text{ h}$ or about 36.5 kWh

To see how much CO_2 this corresponds to, we have to consider what proportion of the electricity in a country is generated from coal (producing a lot of CO_2), natural gas, nuclear energy, renewable sources etc. and the CO_2 emissions associated with each of these methods of generation. A reasonable estimate of CO_2 equivalents per kWh for the US would be about 700 g CO_2e .

So the carbon footprint associated with watching television for one hour a day for a year could be about 36.5×700 , i.e. 26 000 g – or 26 kg of CO_2e .

This calculation considers only the energy associated with using the television and does not take into account the carbon footprint associated with production of the television, its delivery and installation, its disposal etc.

This depends on the size and the type of the television.

The kilowatt-hour is a unit of energy often used for electricity usage.

Nature of science

The problem of dwindling fossil fuel resources is one that transcends any particular science – scientists and technologists from many different disciplines will have to work together to try to solve some of the problems that we will be facing very soon. The problems are, however, not just scientific and scientists will also have to collaborate with politicians, economists, environmental agencies etc. to plan for a world without fossil fuels.

? Test yourself

- 7 Calculate the mass of CO_2 produced per gram of fuel burned and per kJ of energy released when each of the following fuels is burned:
 - a ethanol:
enthalpy change of combustion = $-1367 \text{ kJ mol}^{-1}$
 - b hexane:
enthalpy change of combustion = $-4163 \text{ kJ mol}^{-1}$
- 8 Work out the carbon footprint for each of the following car journeys (assume that the fuel is octane, C_8H_{18} , with a density of 0.703 g cm^{-3}):
 - a 250 km in a car that uses 8.00 dm^3 of fuel per 100 km.
 - b from New York to Chicago (1145 km) in a car that averages 10.00 dm^3 of fuel per 100 km.

C3 Nuclear fusion and fission

This section deals with nuclear reactions – nuclear reactions are very different to the chemical reactions that we have met so far.

Nuclear reactions involve changes in the nuclei of atoms. In these changes, nuclei may:

- give off particles (radiation)
- absorb other particles (such as neutrons)
- split into smaller nuclei (nuclear fission)
- join together to form larger nuclei (nuclear fusion).

During a nuclear reaction, an atom of a particular element may become an atom of a different element.

Contrast this with chemical reactions – these involve valence (outer-shell) electrons. These may be transferred, as in redox reactions or the formation of ions, or shared, as in the formation of covalent bonds. But no chemical reaction involves an atom becoming a different one.

Learning objectives

- Understand that light nuclei can undergo nuclear fusion reactions because this increases the binding energy per nucleon
- Understand that heavy nuclei can undergo nuclear fission reactions because this increases the binding energy per nucleon

C3.1 Nuclear binding energy

The particles in a nucleus (**nucleons**) are held together by very strong forces and energy is released when protons and neutrons come together to form the nucleus. The nuclear **binding energy** is the energy required to break apart the nucleus into protons and neutrons again. The binding energy is not something that the nucleus ‘has’ – it is the energy released when the nucleus is formed or that required to break it apart again. If the binding energy for a nucleus is divided by the total number of nucleons we get the average binding energy per nucleon which is a measure of the stability of the nucleus.

The greater the binding energy per nucleon, the more stable the nucleus.

Figure C.6 shows a graph of binding energy per nucleon against nucleon number (mass number).

The most stable nuclei have mass numbers around 60 (shaded in pink in Figure C.6) – these have the highest binding energy per nucleon. The most stable nuclei are ^{56}Fe and ^{62}Ni . The arrows on the graph in Figure C.6 indicate the type of nuclear processes that are likely to occur.

Heavier nuclei (to the right of the shaded region in Figure C.6) can undergo fission reactions, in which they split up into smaller nuclei, which increase the binding energy per nucleon. Lighter nuclei (to the left of the shaded region in Figure C.6) can undergo fusion reactions, in which they join together to form heavier nuclei, giving a higher binding energy per nucleon.

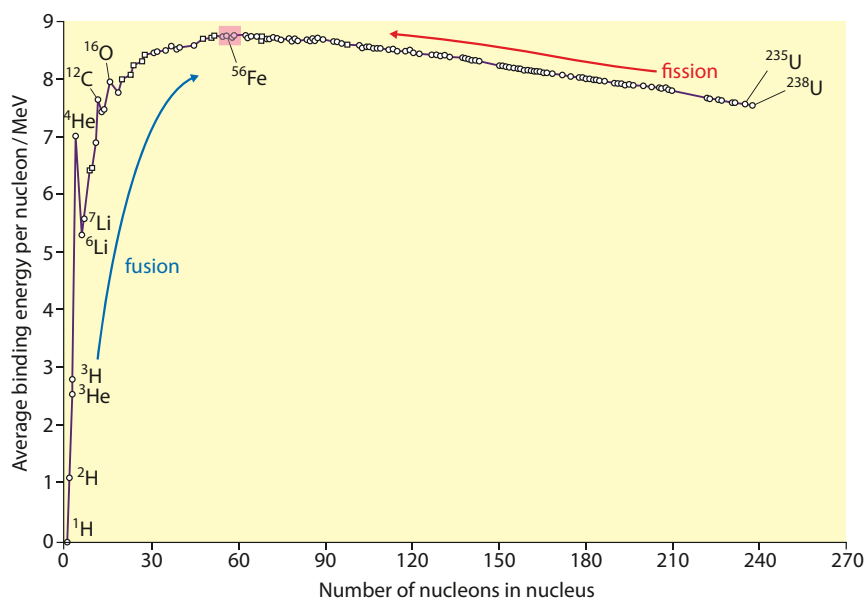
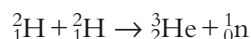


Figure C.6 Average binding energy per nucleon varies with the total number of nucleons.

C3.2 Nuclear fusion

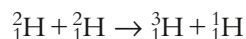
Nuclear fusion is the joining together of smaller nuclei to make a larger one.

A typical fusion reaction is the joining of two deuterium nuclei to form a helium nucleus:



It can be seen from Figure C.6 that the average binding energy per nucleon is greater for ${}^3_2\text{He}$ than for ${}^2_1\text{H}$ and that therefore energy is released in this process (in the form of kinetic energy of the particles, which will manifest itself as heat). Note that the mass numbers and atomic numbers must balance in these equations.

Another possible fusion reaction involving deuterium is:



Although ${}^1_1\text{H}$ has a lower binding energy per nucleon than ${}^2_1\text{H}$, ${}^3_1\text{H}$ has a significantly higher binding energy per nucleon and so, overall, there is an increase in the binding energy per nucleon and energy is released.

Nuclear fusion is seen as a potential energy source of the future. It has many advantages over nuclear fission (discussed below) and the burning of fossil fuels:

- It has the potential to produce vast amounts of energy – 1 g of deuterium has the potential to release hundreds of thousands of times the amount of energy released by burning 1 g of coal.
- It does not produce greenhouse gases (but greenhouse gases may result from the production of deuterium etc.).
- It does not produce radioactive waste.

Learning objectives

- Write equations for nuclear fusion reactions
- Understand nuclear fusion reactions in terms of binding energy per nucleon
- Understand that nuclear fusion is potentially a very important energy source
- Understand that absorption spectra can be used to analyse the elements present in stars
- Understand how an absorption spectrum is formed

${}^1_0\text{n}$ represents a neutron.

Nuclear fusion reactions occur in our Sun and therefore nuclear fusion reactions could be regarded as the ultimate source of all energy on Earth.

- The fuel (deuterium) is abundant – its natural abundance is about 0.015%, which corresponds to about 2×10^{16} kg of deuterium in the world's oceans.
- Deuterium is relatively cheap although several factors must be considered here – for example, its production by electrolysis requires a large input of energy.

There is, however, one major disadvantage of nuclear fusion – there is not one working nuclear fusion power plant. No one has yet managed to produce and contain a sustained nuclear fusion reaction that generates more energy than needs to be put in to produce the extremely high temperatures required.

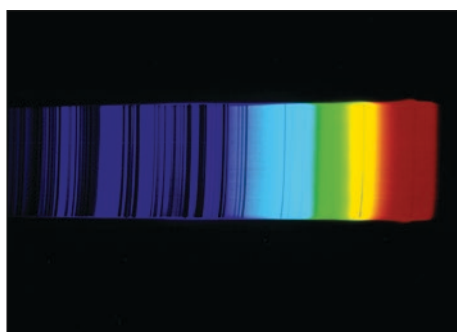


Figure C.7 Solar spectrum showing absorption lines.

When an electron returns to a lower energy level, light is given out – because this is emitted in all directions, very little reaches the observer and so this frequency is essentially missing from the spectrum.

Atomic absorption spectra

Nuclear fusion reactions occur in stars such as our Sun. The Sun emits a continuous spectrum in the visible region but some frequencies are seen to be missing from the spectrum when it is analysed using a spectroscope (Figure C.7).

The dark lines, often called Fraunhofer lines, arise because electromagnetic radiation is absorbed by cooler gases in the outer atmosphere of the Sun. Energy, in the form of certain frequencies of electromagnetic radiation, is absorbed to promote electrons from lower energy levels to higher ones (Figure C.8) in atoms. This frequency of light will therefore be missing from the spectrum as observed by someone on Earth.

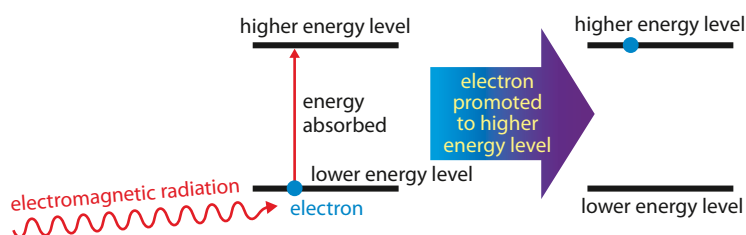


Figure C.8 Absorption of electromagnetic radiation by an atom.

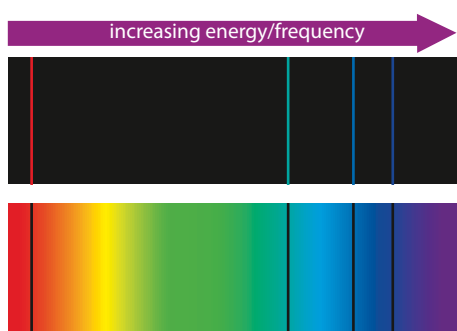


Figure C.9 The relationship between the emission (top) and absorption (bottom) spectra of hydrogen.

An absorption spectrum can also be generated in the laboratory by passing white light (all frequencies) through a sample of gas and analysing the light emerging using a spectroscope.

Each element has a characteristic absorption spectrum – the absorption spectrum of hydrogen is completely different to the absorption spectrum of helium, which is completely different to the absorption spectrum of sodium etc. The dark lines in an absorption spectrum have the same frequency as the lines in the atomic emission spectrum of an element (Figure C.9). Each line in the emission spectrum is caused by an electron falling from a higher atomic energy level to a lower one, whereas each line in the absorption spectrum corresponds to the energy absorbed when an electron is promoted from a lower to a higher energy level. Because the energy levels in the atom are the same, the lines occur at the same frequency in emission and absorption spectra.

The lines in the emission spectrum of hydrogen shown in Figure C.9 are caused by electrons falling from a higher energy level to level 2, whereas the dark lines in the absorption spectrum are caused by energy being absorbed to promote electrons from level 2 to a higher energy level.

The absorption spectrum of helium is much more difficult to analyse than that of hydrogen. The simple model of atomic energy levels can no longer be applied due to repulsion between the two electrons meaning that many different energy states exist.

The elements present in stars can be determined from analysis of the absorption lines in their spectra. The presence of an absorption at a particular frequency, which also corresponds to a line in the emission spectrum of an element, indicates the presence of that element (Figure C.10).

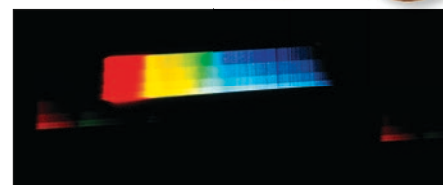


Figure C.10 The spectrum of the star Arcturus showing absorption lines.

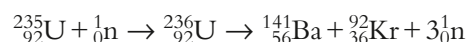
C3.3 Nuclear fission

Nuclear fission

Nuclear fission is the breakdown of a larger nucleus into two smaller fragments of comparable masses.

The average binding energy per nucleon of the daughter nuclei (products of the reaction) is higher than that of the parent nucleus, therefore energy is released in the fission process.

The main fuel used in nuclear reactors is ^{235}U . When bombarded with neutrons, a ^{235}U nucleus absorbs one neutron to produce the highly unstable ^{236}U , which almost instantaneously (picoseconds) breaks apart into two smaller nuclei and more neutrons. A typical reaction is:



The total numbers of protons and neutrons do not change in this type of reaction and so

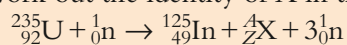
the total mass number and total atomic number on each side must balance.

Learning objectives

- Understand how to write an equation for the fission of uranium-235
- Understand what is meant by critical mass
- Understand that plutonium-239 can be produced in a breeder reactor
- Understand some of the problems associated with nuclear power
- Understand that nuclear waste may contain isotopes with short or long half-lives
- Understand some of the problems associated with the storage and disposal of nuclear waste

Worked example

C.2 Work out the identity of X in the nuclear equation:



The atomic numbers and mass numbers must balance on both sides of the equation.

Atomic number: the total atomic number on the left-hand side is 92, therefore the total atomic number on the right-hand side must also be 92.

$$Z + 49 = 92, \text{ and so } Z = 43$$

From the periodic table we can work out that X is technetium.

Mass number: the total mass number on the left-hand side is $235 + 1 = 236$, therefore the total mass number on the right-hand side must also be 236.

$$125 + A + (3 \times 1) = 236, \text{ and so } A = 108$$

X is technetium-108 or $^{108}_{43}\text{Tc}$.

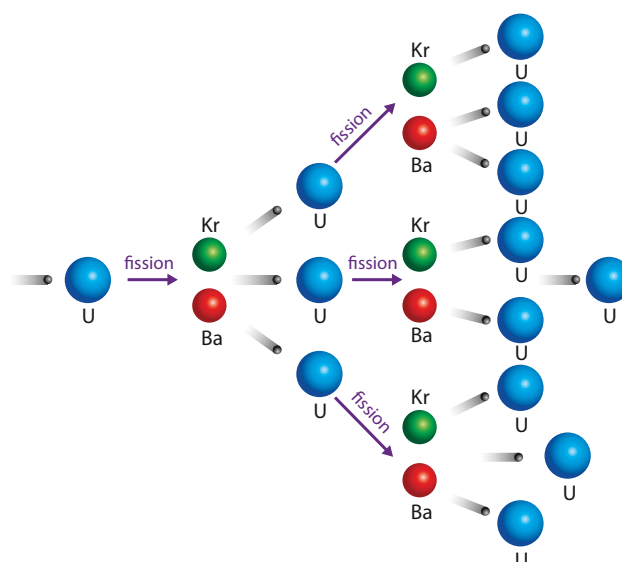


Figure C.11 A nuclear fission chain reaction.

In a nuclear reactor, the neutrons released in the fission process can cause fission of other nuclei, which can result in a chain reaction (Figure C.11). Some of the neutrons produced in a fission reaction will escape from the uranium sample without causing further fission. If all the neutrons escape then a chain reaction will not occur and the sample of uranium is said to be subcritical (Figure C.12a). The mass of uranium needed in order for the reaction to become self-sustaining is called the **critical mass**. For a chain reaction to be self-sustaining, at least one neutron produced in each fission process must go on to cause another fission reaction (Figure C.12b). At the critical mass exactly one neutron from each fission process goes on to cause another fission process.

In a nuclear reactor there must be a sufficient mass of uranium present so that enough neutrons cause fission before escaping from the reactor core. Control rods are inserted between the fuel rods to absorb neutrons in order to keep the average number of neutrons from each fission process that go on to cause further fission at about one. If too many neutrons cause fission the chain reaction could go out of control with potentially catastrophic consequences.

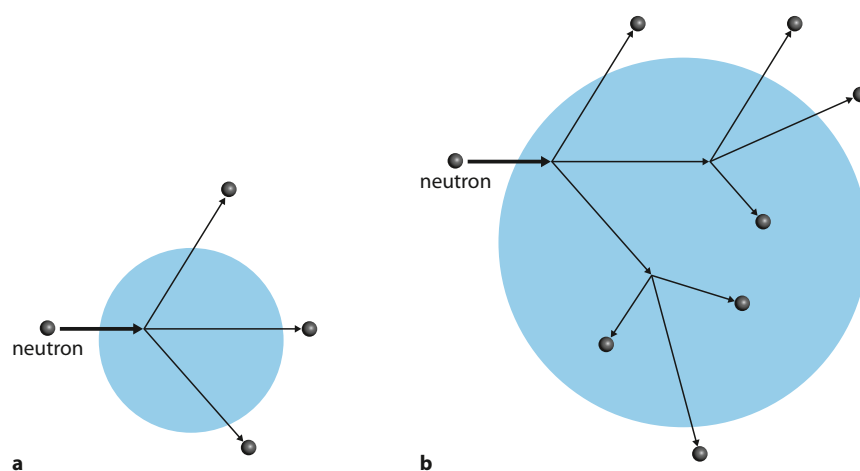
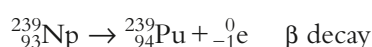
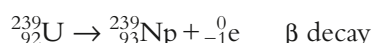


Figure C.12 **a** A subcritical sample of uranium – the neutrons escape before causing further fission; **b** a sample of uranium larger than the critical mass (supercritical) – a self-sustaining chain reaction is produced.

Breeder reactors

The fuel used in a conventional nuclear reactor is uranium enriched with ^{235}U . Naturally occurring uranium contains a very low proportion of ^{235}U (natural abundance 0.72%) – it is mostly ^{238}U , which does not undergo fission. **Breeder reactors** were developed to increase the amount of fissionable nuclear fuel available. ^{235}U undergoes fission most readily with slower-moving neutrons – the neutrons produced in fission processes are slowed down by a moderator so that they are more readily absorbed by ^{235}U nuclei. However, in a breeder reactor there is no moderator and fast-moving neutrons are more readily captured by ^{238}U nuclei which then undergo a series of nuclear reactions to produce ^{239}Pu , a fissionable fuel:



Most breeder reactors use ^{239}Pu as the fissionable fuel – and through the above reactions, starting with uranium-238, they can be designed to produce more fuel than they consume.

Problems associated with using nuclear power

There are many problems associated with the generation of electricity using nuclear fission.

- **Health issues** – there are severe risks associated with exposure to radioactive materials. Exposure to radioactivity can cause radiation poisoning (radiation sickness), cancer and birth defects in children. The workers in nuclear power stations are protected from exposure and the radioactivity level is monitored but escape of radioactive material resulting from an accident poses a severe risk to the health of people living nearby. There have been several examples of nuclear accidents causing a release of radioactive material into the environment – the two most recent are Chernobyl (1986) and Fukushima Daiichi (2011). There is an extensive debate as to how many people died as a result of the Chernobyl disaster and estimates range from hundreds to hundreds of thousands depending on the political/industrial associations of the organisations producing the data.
- **Meltdown** – this is a term usually used to describe accidents in a nuclear reactor. Heat is generated by fission in a nuclear reactor and this is removed from the core by a coolant so that the core remains at a fairly constant temperature. If the nuclear chain reaction is allowed to go out of control or there is a problem with the coolant then the core can overheat. If this is not dealt with, the materials in the core can begin to melt. This can result in radioactive material escaping into the environment. The Chernobyl and Fukushima Daiichi disasters both involved core meltdown.
- **Nuclear weapons** – nuclear reactors, especially breeder reactors, produce ^{239}Pu . Reprocessing of spent fuel from nuclear power stations can produce weapons-grade plutonium. There are severe worries that terrorists could get hold of this and build nuclear weapons. There are hundreds of tonnes of civil-controlled (non-military) plutonium stored around the world – only a few kilograms are needed to make a bomb!

A breeder reactor is a nuclear reactor that produces more fissionable material than it consumes.

β decay involves a neutron in a nucleus becoming a proton and an electron; the electron is ejected from the nucleus.



Breeder reactors have severe reliability problems, use liquid sodium as a coolant (there are dangers associated with this) and are expensive. These reasons, among others, mean that breeder reactors have not generally been a commercial success and many countries have discontinued programs to use them. There are also very great security concerns about the fact that reprocessing of the fuel from breeder reactors can produce large amounts of weapons-grade plutonium.



The plutonium produced by reprocessing is controlled by just a few countries around the world. The UK is believed to have the largest stock of non-military plutonium.

Nuclear waste

Radioactive waste is produced in nuclear power stations, research laboratories, military establishments, industry and hospitals. Numerous radioisotopes are used or produced in various processes and these have half-lives that vary enormously. Some have very short half-lives – ^{131}I , used as a tracer in medicine, has a half-life of just eight days. Others, such as the ^{239}Pu produced in nuclear power stations, have half-lives of thousands of years.

Radioactive waste can be divided into different categories: low-level and high-level. The category ‘intermediate-level’ is also sometimes used. The classification of radioactive waste is important in determining how it can be disposed of and the safety measures that must be used in its transport and handling. The criteria used for the classification of radioactive waste are quite complex and low-level waste is divided into sub-categories (A, B, C and >C) depending on its activity. Generally, low-level waste has lower activity and usually contains isotopes with short half-lives.

Low-level waste

This includes items that have been contaminated with radioactive material or exposed to radioactivity. Examples are gloves, protective clothing, tools, soil, rubble and carcasses of animals that have been treated with radioactive materials. Low-level waste may be stored on site until it has decayed to such an extent that it can be disposed of as ordinary waste (e.g. in landfill sites or released into the sewage system) or shipped to a central site for more specialised disposal. Some low-level waste is incinerated, which reduces its volume considerably and distributes the radioisotopes over a wide area. The ash from incineration is assessed for activity and disposed of appropriately. Low-level waste with higher activity is often just buried underground (‘near-surface’ disposal) – for example, in individual concrete canisters or in concrete-lined vaults. Low-level waste may need to be contained underground for up to 500 years depending on its activity and half-life.

High-level waste

This includes spent fuel rods or other materials resulting from the reprocessing of nuclear fuels. It contains fission products and transuranic (beyond uranium) elements generated in the reactor core. This spent fuel has a high concentration of radioactive isotopes, some of which (such as ^{239}Pu) will remain hazardous to humans and other living things for thousands of years. Spent fuel rods may either be considered as waste and disposed of directly, or reprocessed to extract material to be used as more nuclear fuel. Reprocessing of nuclear fuel produces a highly radioactive liquid waste that can be converted to glass (vitrified) to make storage easier.

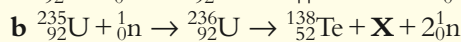
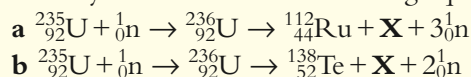
High-level waste is initially stored on site at nuclear power plants in storage pools (cooling ponds). Storage under water is usually for a minimum of nine months but sometimes the spent fuel rods are stored in this way for decades. After sufficient cooling, the fuel rods may be transported to a reprocessing plant, left in the pools or transferred to dry storage casks. These dry casks have very thick walls and are made of steel and concrete. The dry casks are then stored in concrete bunkers. Most spent fuel rods from nuclear power stations in the US are left in cooling ponds or transferred to dry storage awaiting a more permanent method of storage.

Permanent storage of high-level radioactive waste is a major problem and several solutions, such as burying the waste deep underground in stable geological areas, have been suggested. Over thousands of years, however, it is difficult to envisage what processes could occur that might lead to the release of the radioactive material.

? Test yourself

- 9 Write balanced equations for each of the following fission reactions of uranium-235:
- to form caesium-140 and rubidium-93
 - to form molybdenum-105 and tin-129

10 Identify X in each of the following equations



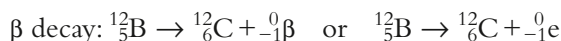
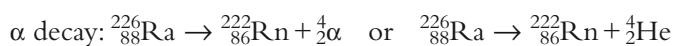
C3.4 Half-life

Radioactive decay involves changes in the nucleus of an atom resulting in particles – and sometimes electromagnetic radiation – being emitted from the nucleus. The two main particles that are emitted are:

- alpha particles, ${}^4_2\alpha$, which are helium nuclei composed of two protons and two neutrons
- beta particles, ${}^0_{-1}\beta$, which are electrons formed when a neutron turns into a proton and an electron.

Emission of γ rays (high-energy electromagnetic radiation) may also accompany radioactive decay.

Examples of radioactive decay processes are:



Radioactive decay is a random process and it is impossible to predict when any one particular nucleus will decay but on average a sample of any radioisotope with a very large number of atoms will decay with constant half-life.

Half-life is the time it takes for the number of radioactive nuclei present in a sample at any given time to fall to half its value.

Half-life ($t_{1/2}$) varies from isotope to isotope – for example the half-life of ${}^{226}\text{Ra}$ is 1600 years but that of ${}^{224}\text{Ra}$ is 3.7 days. Half-life is independent of the ambient conditions (temperature, pressure etc.) and also of the mass of a radioactive sample – the half-life is the same whether 1 g or 1 kg of a particular isotope is present.

Figure C.13 shows a graph of the decay of an isotope with a half-life of 2 s – the number of original nuclei remaining decreases by half every 2 s.

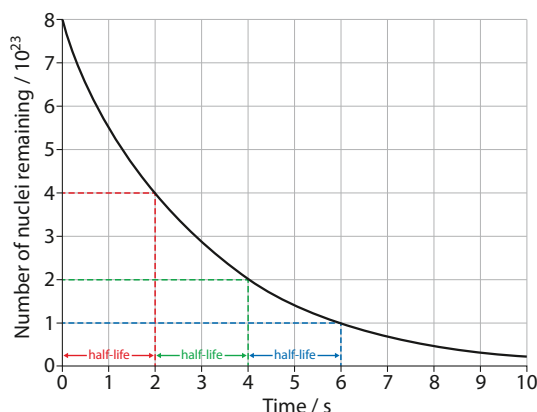


Figure C.13 Decay of a radioisotope with half-life of 2 s – this is exponential decay.

Learning objectives

- Understand what is meant by half-life
- Solve problems involving integral numbers of half-lives

Radium-226 has a half-life of approximately 1600 years. If we start with 1 g of pure ^{226}Ra , after 1600 years there will be 0.5 g left, after a further 1600 years there will be 0.25 g left and after a total of 4800 years (three half-lives) there will be only 0.125 g of radium-226 left:

$$1\text{ g} \xrightarrow[1600\text{ years}]{\text{half-life}} 0.5\text{ g} \xrightarrow[1600\text{ years}]{\text{half-life}} 0.25\text{ g} \xrightarrow[1600\text{ years}]{\text{half-life}} 0.125\text{ g}$$

Half-life can also be expressed in terms of the activity of a sample – this is the number of nuclei that decay per second. The half-life is then the time taken for the activity to drop to half its original value.

Worked examples

C.3 Germanium-71 has a half-life of 11 days. If there were originally 2.00 mg of this isotope present in a sample, calculate the mass remaining after 44 days.

It takes 11 days for the amount to fall by half, therefore there will be 1.00 mg present after 11 days. This will drop to 0.500 mg after another 11 days, 0.250 mg after another 11 days and 0.125 mg after a further 11 days:

$$2.00\text{ mg} \xrightarrow[11\text{ days}]{\text{half-life}} 1.00\text{ mg} \xrightarrow[11\text{ days}]{\text{half-life}} 0.500\text{ mg} \xrightarrow[11\text{ days}]{\text{half-life}} 0.250\text{ mg} \xrightarrow[11\text{ days}]{\text{half-life}} 0.125\text{ mg}$$

Therefore the mass of germanium-71 remaining will be 0.125 mg.

C.4 The half-life of uranium-238 is 4.5×10^9 years. Calculate how long it would take 32 g of uranium-238 to decay to 1 g.

This decay involves five half-lives:

$$32\text{ g} \xrightarrow{\text{half-life}} 16\text{ g} \xrightarrow{\text{half-life}} 8\text{ g} \xrightarrow{\text{half-life}} 4\text{ g} \xrightarrow{\text{half-life}} 2\text{ g} \xrightarrow{\text{half-life}} 1\text{ g}$$

So the total time is $5 \times 4.5 \times 10^9 = 2.3 \times 10^{10}$ years.

C.5 Calculate the half-life of protactinium-233 if it takes 108 days for 100 mg of the element to decay to 6.25 mg.

This decay will take four half-lives:

$$100\text{ mg} \xrightarrow{\text{half-life}} 50\text{ mg} \xrightarrow{\text{half-life}} 25\text{ mg} \xrightarrow{\text{half-life}} 12.5\text{ mg} \xrightarrow{\text{half-life}} 6.25\text{ mg}$$

So 108 days is equivalent to 4 half-lives. 1 half-life is therefore $\frac{108}{4} = 27$ days.

C.6 Calculate the time taken for the activity of a sample of rubidium-83 to fall to 12.5% of its original value given that its half-life is 86 days.

The data here are given in terms of activity instead of mass, but the answer is worked out in the same way:

$$100\% \xrightarrow[86\text{ days}]{\text{half-life}} 50\% \xrightarrow[86\text{ days}]{\text{half-life}} 25\% \xrightarrow[86\text{ days}]{\text{half-life}} 12.5\%$$

Three half-lives are required for this decay and so the total time is $3 \times 86 = 258$ days.



? Test yourself

- 11 In each of the following questions, calculate the amount remaining from a 100 mg sample of the given radioisotope after the specified time.
- ^{105}Rh has a half-life of 35 h. Calculate the mass remaining after 70 h.
 - ^{209}Po has a half-life of 105 y. Calculate the mass remaining after 420 y.
 - ^{219}Rn has a half-life of 3.96 s. Calculate the mass remaining after 39.6 s.
- 12 Calculate the half-lives of each of the following radioisotopes.
- It takes 180 days for 80 mg of iron-59 to decay to 5 mg.
 - It takes 2.1×10^{12} years for 60 mg of platinum-190 to decay to 7.5 mg.
- 13 Calculate how long each of the following decay processes will take:
- 32 mg of silicon-32 (half-life 160 y) to decay to 1 mg
 - 56 mg of mendelevium-258 (half-life 56 d) to decay to 7 mg.

Nature of science

There are many ethical issues associated with the development of nuclear energy and nuclear weapons. Arguments in favour of nuclear power include the fact that it does not produce greenhouse gases and arguments against focus on the disposal of nuclear waste, safety and the potential to be used in bombs.

The play *Die Physiker* by Friedrich Dürrenmatt provides an interesting treatment of the responsibilities of scientists towards the wider community.

The atomic bombs dropped on Japan in 1945 used fission reactions, the hydrogen bomb also involves a fusion reaction.

C4 Solar energy

Most forms of energy that we use come originally from the Sun – for example, coal is derived from plants that used photosynthesis to store chemical energy and wind power comes from differential heating of the Earth's surface by the Sun. The only form of energy that does not come from the Sun is nuclear – although the uranium was, of course, made in a sun (star) by nuclear fusion reactions. The Sun's energy comes from nuclear fusion reactions that involve the joining together of small nuclei, such as hydrogen, with the release of a great deal of energy.

C4.1 Coloured compounds and photosynthesis

Coloured compounds

Chromophores

In order to absorb electromagnetic radiation in the ultraviolet and visible (UV–Vis) regions of the electromagnetic spectrum, molecules must generally contain a double bond in the form of $\text{C}=\text{C}$, $\text{C}=\text{O}$ or a benzene ring. These groups, which give rise to absorptions in the UV–Vis region, are called **chromophores**.

Learning objectives

- Understand what is meant by a conjugated system
- Understand that molecules with a sufficiently long conjugated system can absorb visible light
- Understand that photosynthesis is the process by which plants turn light energy into visible light
- Write an equation for photosynthesis

A **conjugated system** is a system of alternating single and double bonds in a molecule.

Conjugated systems

The double bonds highlighted in Figure C.14 form a conjugated system but the two double bonds at the ends of the molecule are not part of this system as they are separated from the other double bonds by more than one single bond.

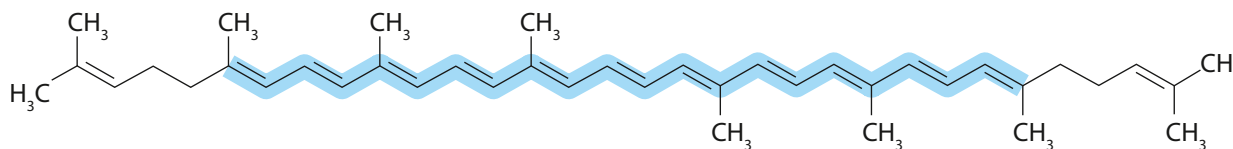


Figure C.14 Lycopene, the red pigment in tomatoes, has 11 conjugated double bonds. Note that not all the C=C units in lycopene are part of the conjugated system.

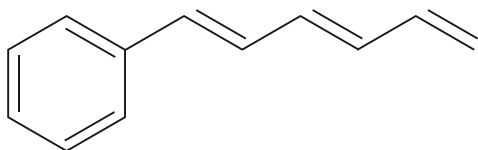


Figure C.15 A molecule with six conjugated double bonds.

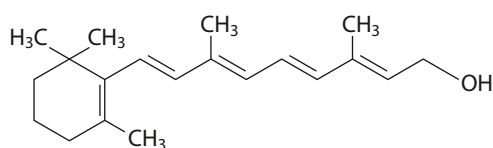


Figure C.16 Retinol has only five conjugated double bonds and absorbs only in the UV region of the electromagnetic spectrum.

The double bonds must alternate with single bonds for a system to be conjugated – if there are two or more single bonds between the double bonds then the system is not conjugated.

We have already seen that a benzene ring can be represented as a ring with three alternating double and single bonds – so a benzene ring is a conjugated system. The molecule shown in Figure C.15 has six conjugated double bonds.

Absorption of electromagnetic radiation and colour

In order for a compound to be coloured, its molecules must absorb visible light. Visible light is electromagnetic radiation with wavelengths between about 400 and 750 nm, so if a molecule absorbs radiation between these wavelengths it will be coloured. The longer the conjugated system, the longer the wavelength of the radiation absorbed and if a conjugated system involves more than about eight double bonds, the molecules should absorb in the visible region of the spectrum and be coloured.

Lycopene has a system of 11 conjugated double bonds (Figure C.14) and absorbs light in the blue–green part of the visible spectrum, and therefore appears red. Retinol (see Figure C.16), however, only has a system of five conjugated double bonds and therefore does not absorb visible light (it only absorbs ultraviolet radiation) and is colourless.

Chlorophyll a and b have long conjugated systems (highlighted in Figure C.17). They absorb light in the 400–500 nm region and in the 600–700 nm region (Figure C.18). The green light in the middle part of the spectrum is not absorbed, and so chlorophyll appears green.

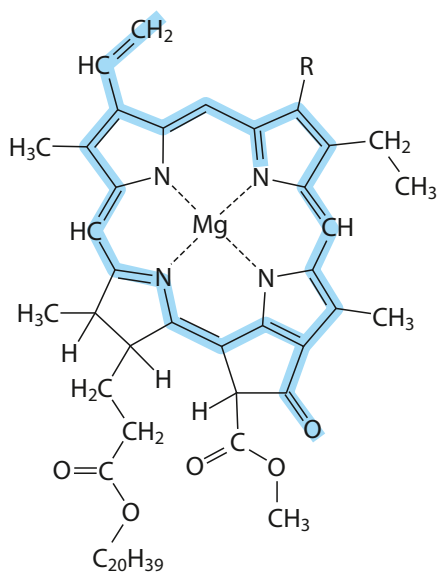


Figure C.17 The basic structures of chlorophyll a (in which R is CH₃) and chlorophyll b (R is CHO) showing the conjugated system. Chlorophyll b has an extra double bond that is part of the conjugated system.

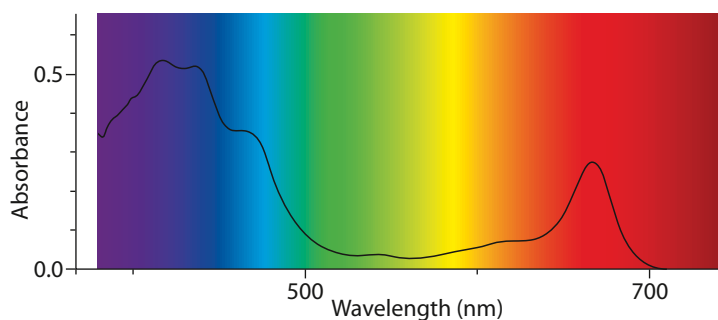


Figure C.18 The visible spectrum of chlorophyll.

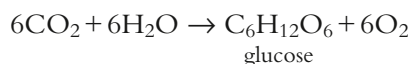


Photosynthesis

The substances needed for photosynthesis are:

- sunlight – the red part of the light is absorbed by plants
- carbon dioxide – from the air
- water – from the air and the ground
- chlorophyll – a green pigment in plants, which absorbs red light.

The exact nature of the reactions which occur is complex but the basic equation is:



The energy absorbed by the chlorophyll is used to drive this process; it is endothermic.

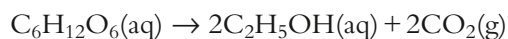
Plants use the glucose produced as an energy source to live and grow (the energy is ultimately released again in the process of respiration, the reverse of the above reaction).

Photosynthesis involves the conversion of light energy from the Sun to chemical energy.

C4.2 Biofuels

Fermentation

The basic reaction that happens during fermentation is the enzyme-catalysed conversion of glucose to ethanol:



Yeast, which contains enzymes, is mixed with a solution of glucose and the reaction mixture kept at a temperature of about 30 °C. Fermentation produces a mixture with an ethanol concentration of about 8–12% and this mixture must be distilled to concentrate the ethanol.

Ethanol produced in this way is a **biofuel** and is mixed with gasoline for use as a fuel in cars. Various blends of ethanol with gasoline are available and, especially in Brazil and the US, flexible-fuel vehicles are available that have been designed to work with all possible mixtures of ethanol and gasoline.

The main producers of ethanol as a fuel by fermentation are Brazil and the US. The main feedstock for fermentation in Brazil is sugar cane, whereas in the US it is corn starch. A disadvantage with corn starch is that it must be pre-treated, which involves heating and mixing with enzymes to produce simple sugars that can undergo fermentation.

Ethanol has an octane number of 111 and so it increases the octane number when added to most fuels.

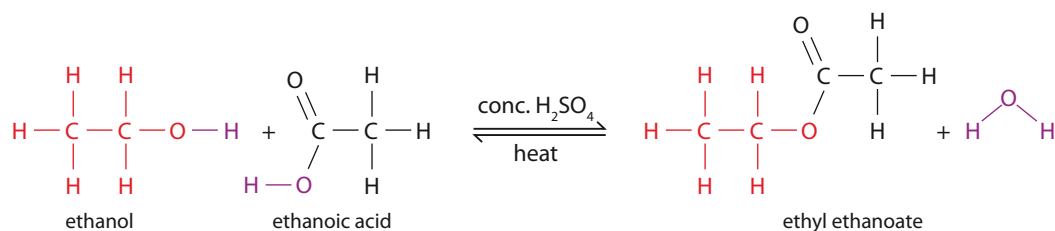
Learning objectives

- Understand that fermentation can be used to produce ethanol for use as a biofuel
- Write an equation for fermentation of glucose
- Understand what is meant by a transesterification reaction
- Write equations for transesterification reactions
- Understand that vegetable oils are too viscous to be used in modern diesel engines
- Understand that biodiesel can be made using a transesterification reaction
- Understand why biodiesel has a lower viscosity than vegetable oils
- Understand some of the advantages and disadvantages of using biofuels

A biofuel is a fuel produced from organic matter obtained from plants, waste material etc.

Transesterification reactions

Esters contain the COOC functional group and are made in the reaction between a carboxylic acid and an alcohol (see Topic 10):



Exam tip

To work out the structure of the new ester formed, the alkyl group of the original ester (on the C-O side) is replaced by the alkyl group of the alcohol that it reacts with.

In a transesterification reaction an ester is reacted with a different alcohol which substitutes for the original one (Figure C.19). This reaction occurs in the presence of a strong acid or base catalyst such as sulfuric acid or sodium hydroxide.

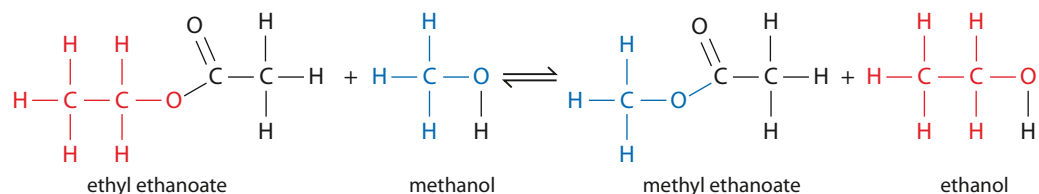
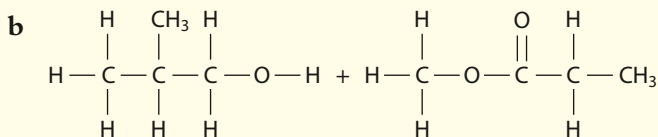
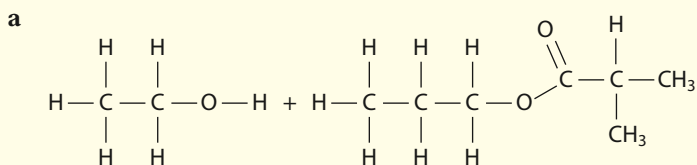


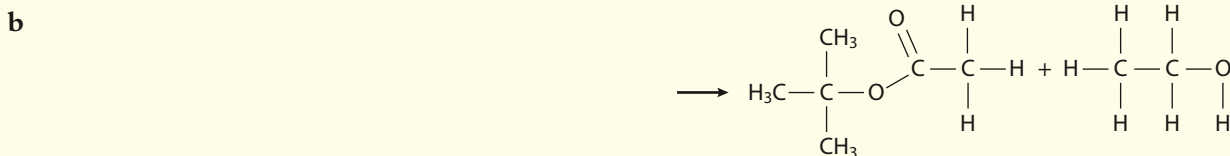
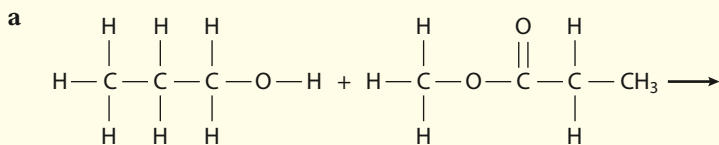
Figure C.19 A transesterification reaction.

? Test yourself

14 Predict the structure of the ester formed in each of the following transesterification reactions.



15 Copy and complete the equations for these transesterification reactions using structural formulas.



Biodiesel

The specific energy of peanut oil is about $37\,000\text{ kJ kg}^{-1}$, whereas its energy density is about $34\,000\text{ kJ dm}^{-3}$. These values can be compared with those for petrodiesel – approximately $43\,000\text{ kJ kg}^{-1}$ and $36\,000\text{ kJ dm}^{-3}$. The original diesel engine, as designed by Rudolf Diesel, ran on peanut oil and it can be seen that the energy released by burning fats and oils is very similar to that by burning diesel fuel. However, vegetable oils cannot be used directly in modern diesel engines because they are too viscous. There are also other disadvantages such as the fact that they can solidify at lower temperatures.

Biodiesel can be made from plant oils. The structure of a plant oil is shown in Figure C.20. The molecule contains three ester linkages, one of which is highlighted, and is formed when three carboxylic acid molecules (fatty acids) react with propane-1,2,3-triol (glycerol). Because three fatty acid units form ester linkages to propane-1,2,3-triol, these molecules are called **triglycerides** (if two fatty acid molecules were joined it would be a **diglyceride**).

Biodiesel can be made when vegetable oils undergo a transesterification reaction – the fatty acid molecules form an ester linkage to a different alcohol (usually methanol). The reaction is usually carried out in the presence of a strong base – the methanol is first of all reacted with solid sodium hydroxide to form sodium methoxide (CH_3ONa), which is then reacted with the vegetable oil. The overall reaction is shown in Figure C.21.

The biodiesel molecule shown in Figure C.21 belongs to the general class of alkyl esters. If ethanol had been used instead of methanol the product would have been $\text{C}_{17}\text{H}_{33}\text{COOCH}_2\text{CH}_3$.

Because this is a reversible reaction, excess alcohol is used to shift the position of equilibrium towards the products.

Vegetable oils contain a small amount of free fatty acids (more if they have been used for cooking) and these free fatty acids can also be converted to esters when they are reacted with alcohol. When the free

Petrodiesel is diesel fuel obtained from crude oil.

The biodiesel molecule shown in Figure C.21 here is sometimes called a fatty acid methyl ester, FAME.

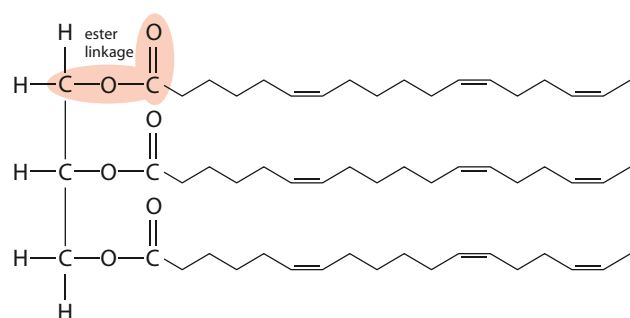


Figure C.20 A vegetable oil.

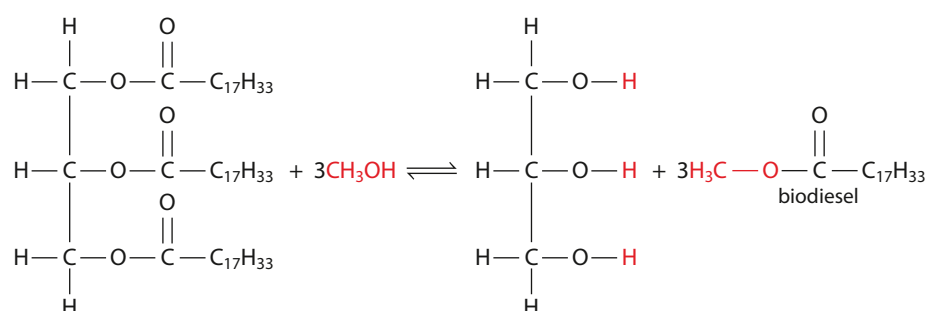


Figure C.21 Biodiesel can be made in a transesterification reaction.

fatty acid content is very high either some sort of pre-treatment is used to reduce their concentration and/or an acid catalyst is used.

The biodiesel must be separated from the glycerol (the glycerol is denser than the biodiesel and forms a separate layer at the bottom) and then purified before it can be used.

Viscosity is a measure of the resistance of a liquid to flow.

Liquids such as honey have a much higher viscosity than water. The viscosity of biodiesel is significantly lower than that of the original oil and is similar to that of petrodiesel. Viscosity depends, among other things, on the strength of intermolecular forces. Biodiesel molecules have much lower relative molecular masses than the oil molecules from which they were made and therefore have weaker London forces between molecules. Because the forces between biodiesel molecules are weaker they have less resistance to flow and a lower viscosity. The lengths of the hydrocarbon chains in biodiesel and petrodiesel are very similar and therefore the London forces between chains will be similar, and so will their viscosities.

Advantages and disadvantages of biofuels

One of the main advantages of using biofuels over fuels derived from petroleum is that biofuels are renewable, whereas fuels from petroleum are non-renewable. An advantage of using biofuels for a country such as Brazil is that they can be produced locally and so there is less reliance on expensive oil imports from other countries. Biofuels can also be produced from waste materials, e.g. biogas from manure or landfill sites or biodiesel from waste oil.

Comparing specific properties of biodiesel fuels and petrodiesel fuels:

- biodiesel is biodegradable and so there is less environmental impact from spillages
- biodiesel contains no sulfur so there are no sulfur dioxide emissions when it is burned – however, use of biodiesel can result in slightly higher emissions of NO_x
- biodiesel is more expensive than petrodiesel
- biodiesel has a lower specific energy than petrodiesel
- biodiesel is a better lubricant than petrodiesel and so can reduce engine wear.

Biofuels are sometimes described as being carbon-neutral in that the carbon atoms in the biofuel originally came from carbon dioxide in the atmosphere and these are replaced when the fuel is burned, so there is no net change in the number of carbon dioxide molecules in the atmosphere. However, the situation is more complicated than this and how environmentally friendly a biofuel such as ethanol is depends on several factors. At first sight it would appear to be carbon-neutral in that the six molecules of carbon dioxide released during the fermentation process and the combustion of ethanol balance exactly with the six molecules of carbon dioxide that were removed from the atmosphere when glucose was formed by photosynthesis. However, several other processes that produce carbon dioxide must be considered:

- the agricultural machinery used to work the land where the corn/sugar cane is grown



- the production of fertilisers for use on the farms
- the transportation of corn/sugar cane
- the processing of corn starch to produce simple sugars
- distillation and so on.

All these processes require energy and, assuming that most of this comes originally from burning fossil fuels, ethanol is far from carbon-neutral. The debate is on-going as to whether some biofuels actually have a *larger* carbon footprint than fuels derived from petroleum. There is also some debate in the US as to whether more energy must be put in to make ethanol from corn than is obtained from it.

Growing biofuel crops requires large areas of land and in a world where there are still severe problems with famine, there are ethical implications in using vast areas of land to grow a foodstuff to be converted into a fuel for vehicles. The use of corn for fuel production has also led to an increase in the cost of this commodity.

Nature of science

It is essential that scientific discoveries that can make a significant difference to society are properly explained to and understood by the public. A basic scientific education is important for all so that they can make sense of headlines such as ‘Brit boffins make petrol from air and water’ – variations on which appeared in many newspapers in the UK in 2012.

C5 Environmental impact – global warming

The greenhouse effect

This is an important mechanism for maintaining the Earth’s temperature at a reasonable level. Without some sort of greenhouse effect the Earth would be too cold to maintain life as we know it.

Some of the short-wavelength solar radiation (visible light) from the Sun that reaches the Earth is reflected back into space, and the rest passes through the atmosphere to reach the Earth’s surface. The surface absorbs some of this radiation and heats up. The warmed surface radiates longer-wavelength infrared radiation. Part of this radiation is absorbed by the greenhouse gases, such as carbon dioxide, in the atmosphere. Some of the radiation absorbed by the greenhouse gases is then re-radiated back to Earth.

The overall effect is that the heat is ‘trapped’ by the gases in the atmosphere (Figure C.22). The natural equilibrium between incoming and outgoing radiation maintains the Earth’s mean temperature at about 15 °C.

If the amount of greenhouse gases in the atmosphere increases then more infrared radiation will be absorbed and re-radiated back to Earth and the global temperature should increase.

Learning objectives

- Understand how the greenhouse effect works
- Understand the mechanism by which greenhouse gases absorb infrared radiation
- Describe the sources of greenhouse gases and the contribution of different gases to global warming
- Understand the evidence for a relationship between increased levels of greenhouse gases and increased global temperature
- Discuss approaches to controlling carbon dioxide emissions
- Explain what is meant by global dimming
- Understand that the carbon dioxide in the atmosphere is in equilibrium with the dissolved carbon dioxide in the oceans
- Understand that increased levels of carbon dioxide in the atmosphere can result in ocean acidification

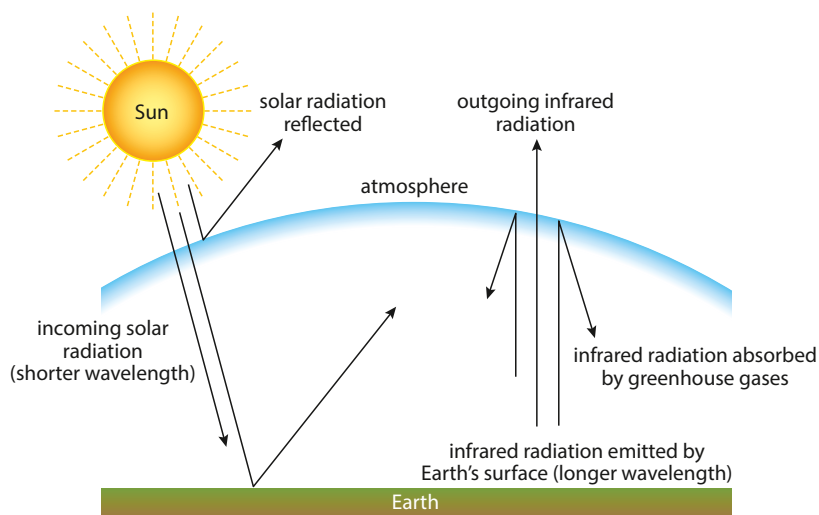


Figure C.22 The greenhouse effect.

The mechanism by which infrared radiation is absorbed by greenhouse gases

Even at absolute zero, atoms in molecules are vibrating relative to each other. Just as the electrons in an atom or molecule can exist only in certain energy levels (the energy of an electron is quantised) the vibrational energy of a molecule is quantised. This means that the vibrational energy of a molecule can take only certain allowed values and not just any value.

For instance, a molecule might be able to exist in either the level with vibrational energy V_1 or that with V_2 (Figure C.23). The molecule can absorb certain frequencies of infrared radiation to move it from a lower vibrational energy level to a higher one.

When molecules of greenhouse gases, such as carbon dioxide in the atmosphere, absorb infrared radiation to promote them to higher vibrational energy levels, they vibrate more energetically. As they move back down to a lower vibrational energy level, the 'extra' energy is given out again. This energy is given out in all directions and so some infrared radiation will be radiated back to Earth.

Diatomic molecules

For a molecule to be able to absorb infrared (IR) radiation, there must be a change in bond polarity (more precisely, dipole moment) when it vibrates. Polar molecules constitute a dipole – two charges separated from each other (see Topic 4, page 130 in the Coursebook).

The size of the dipole moment depends on the magnitude of the partial charges and on the distance between them. The longer the distance between the charges, the higher the dipole moment. So as an H–Cl molecule vibrates, the dipole moment increases and decreases (Figure C.24).

This means that a hydrogen chloride molecule is able to absorb IR radiation and we say that it is *IR-active*.

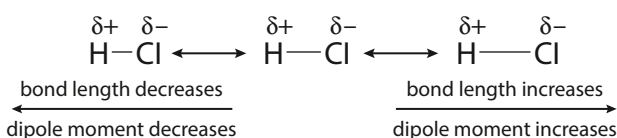


Figure C.24 The vibration of an HCl molecule.

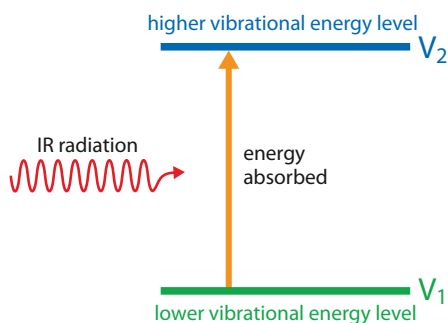


Figure C.23 Molecules can absorb IR radiation to promote them from a lower to a higher vibrational energy level.

The vibrations of a molecule are often likened to two masses joined by a spring.

If infrared radiation is passed through a sample of hydrogen chloride, radiation of wavenumber approximately 2900 cm^{-1} is absorbed. However, oxygen molecules, O_2 , and nitrogen molecules, N_2 , which are the major gases in the Earth's atmosphere, cannot absorb infrared radiation because these molecules are non-polar and their molecular vibrations do not involve a change in dipole moment. O_2 and N_2 are said to be IR-inactive.

A diatomic molecule can absorb infrared radiation only if it is polar.

Triatomic molecules

Triatomic molecules are more complicated than diatomic ones because they can vibrate in different ways. There are two main types of vibrational mode – *stretching* (bond lengths change) and *bending* (bond angles change).

Such molecules can stretch in two different ways – either both bond lengths can increase and decrease together (symmetric stretch) or one gets shorter as the other gets longer (asymmetric stretch). The vibrational modes for a bent triatomic molecule, such as water, are shown in Figure C.25.

The vibrational modes may also be shown as illustrated in Figure C.26. Only vibrational modes that involve a change in dipole moment are IR-active. For polar molecules such as H_2O , all three vibrational modes are IR-active because they all involve changes in the dipole moment. Water molecules therefore absorb infrared radiation at three distinct frequencies, corresponding to the energy to excite the molecule to a higher vibrational energy level for each vibrational mode.

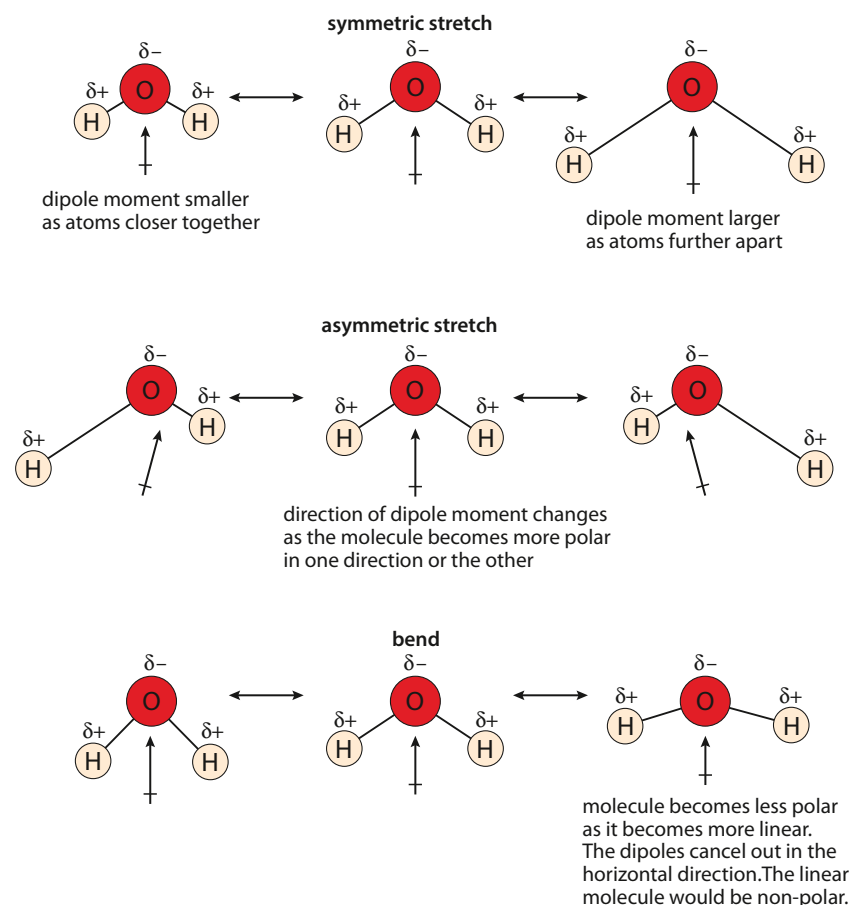


Figure C.25 Vibrational modes of water showing how the dipole changes – all types of vibration are occurring at the same time.

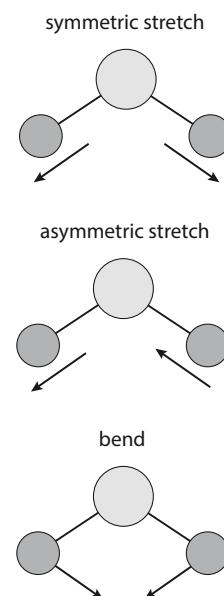


Figure C.26 The types of vibrational mode in a bent triatomic molecule. The central atom can be regarded as fixed in position.

There are actually four different vibrational modes for CO₂ but two are degenerate – they have the same energy.

The vibrational modes of carbon dioxide

Carbon dioxide is a linear non-polar molecule (Figure C.27a). There are three different types of vibrational mode (Figure C.27b) but only two of them (the asymmetric stretch and bend) involve a change in dipole moment and are IR-active. The symmetric stretch is IR-inactive because the molecule remains non-polar when it vibrates in this way (Figure C.27c). Carbon dioxide, therefore, only absorbs infrared radiation at two distinct frequencies.

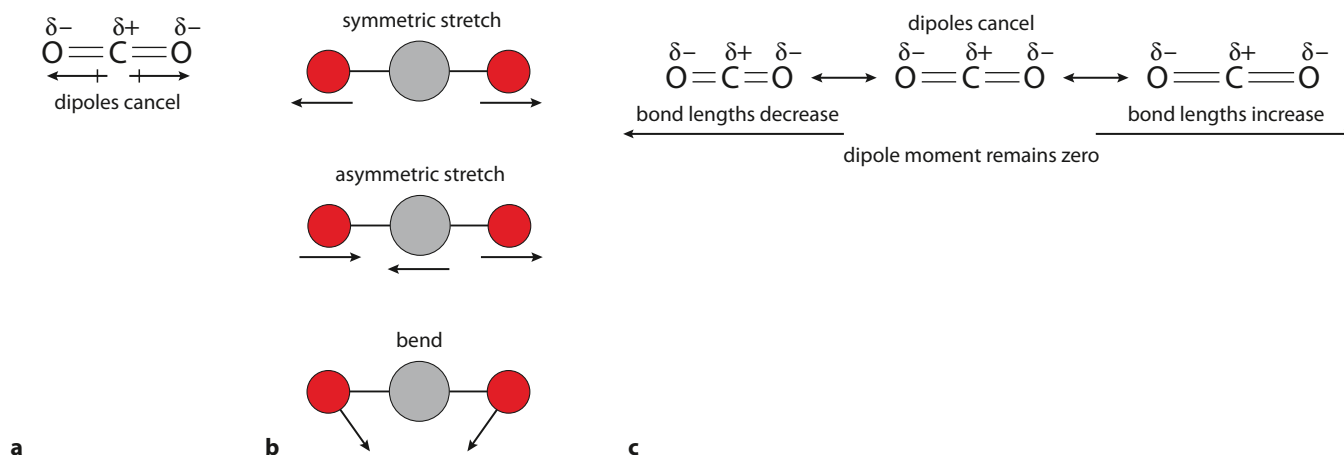


Figure C.27 a The non-polar CO₂ molecule; b the three vibrational modes; c the symmetric stretch which is IR-inactive.

The vibrational modes of methane

The vibrational modes of methane are considerably more complicated – there are four distinct vibrational modes, two of which are IR-active (Figure C.28). Methane is a non-polar molecule and it can be seen that the symmetric stretch, where all four hydrogen atoms move in and out together, involves no change in dipole moment.

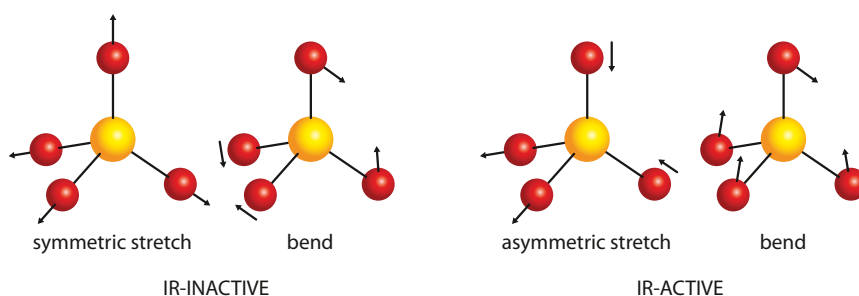


Figure C.28 The vibrational modes of methane.

Extension

Although ozone, O₃, would appear to be a non-polar molecule that does not absorb infrared radiation, an uneven charge distribution in the molecule means that certain vibrational modes are IR-active.

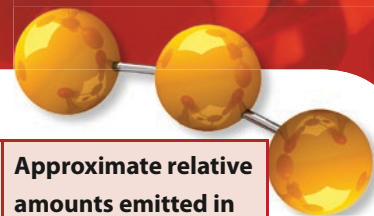
If stretching and bending modes cause a change in dipole moment then a molecule will absorb infrared radiation.



Test yourself

16 Which of the following molecules will absorb infrared radiation?

O₂ HCN HF H₂ CO CO₂ H₂S N₂ CFCI₃ N₂O



Gas	Source	Heat-trapping effectiveness	Approximate relative amounts emitted in US per year
H ₂ O	evaporation from oceans and lakes on the Earth's surface	0.1	
CO ₂	combustion of carbon fuels and biomass*	1	100
CH ₄	anaerobic decay of organic material agriculture: rice fields, marshes, animals (enteric fermentation in the stomachs of ruminant animals such as cows)	26	9.6
N ₂ O	agricultural soil management (fertilisers), nitric acid production	216	5.4
CFCs	refrigerants, propellants, foaming agents, solvents	13 000–23 000	

Table C.8 The main greenhouse gases, their sources and relative heat-trapping ability.

*When considering factors that contribute CO₂ to the atmosphere, we should also consider the effects of removing the mechanisms that reduce the amount of CO₂ in the atmosphere. For instance, if areas of forest are cleared this can increase CO₂ levels in the atmosphere in two ways – first, CO₂ is not being removed by the process of photosynthesis and second, if the wood is burned then CO₂ is produced.

The sources and relative effects of the main greenhouse gases

The main greenhouse gases, their sources and relative heat-trapping ability are shown in Table C.8.

The contribution of a particular greenhouse gas to global warming depends on several factors – its ability to absorb infrared radiation, its abundance in the atmosphere, its atmospheric lifetime and the wavelength range in which it absorbs infrared radiation.

The potential for a particular gas to cause global warming can be described in terms of its global-warming potential. The potential for 1 kg of some gases to cause global warming over a particular time period (e.g. 20 years) is compared to that of 1 kg of CO₂ in Table C.9.

The contribution of each of the anthropogenic gases to global warming between 1980 and 1990 is shown in Table C.10. Methane has a much higher global-warming potential than carbon dioxide, for equal masses, but is produced in much smaller amounts.

Carbon dioxide has a greater influence on global warming than some of the other gases from anthropogenic sources because, even though it does not absorb as much infrared radiation, it is produced in greater amounts.

Water vapour is another important greenhouse gas and most scientists would regard it as the most important greenhouse gas because it is present in the atmosphere at a much higher concentration than other greenhouse gases. However, the amount of water vapour in the atmosphere is only directly influenced to a small extent by human activities.

If the Earth gets hotter through the release of other greenhouse gases, this will increase the evaporation of water and so further increase the amount of water vapour in the atmosphere.

The effect of water vapour on global temperatures is difficult to predict – evaporation of water causes cooling of the Earth's surface (just as evaporation of sweat from your skin causes cooling of your body), but an increase in water vapour in the atmosphere will increase its effect as a greenhouse gas. In addition to this, an increase in water vapour will result in the formation of more clouds which reflect more sunlight back into space and so cause cooling.

Gas	Global-warming potential (20 years)
CO ₂	1
CH ₄	72
N ₂ O	289
CFC-11	3800

Table C.9 The global-warming potential of various gases over 20 years. This compares how effective equal masses of gases are at causing global warming, relative to carbon dioxide.

Gas	Contribution to global warming / %
CO ₂	55
CH ₄	15
N ₂ O	6
CFCs	24

Table C.10 The contribution of anthropogenic gases to global warming between 1980 and 1990.

'Anthropogenic' means that it has been produced by human activities.

The influence of increasing amounts of greenhouse gases on global temperatures

In the 10 000-year period up to 1750, the CO_2 concentration in the atmosphere remained fairly constant at around 280 ppm – this had risen to 379 ppm by 2005. Similarly, CH_4 and N_2O abundances in the atmosphere have increased since the Industrial Revolution.

In the 100 years up to 2005, global temperatures increased by about 0.7°C and in the past 30 years temperatures are estimated to have increased by about 0.2°C per decade. Although there is general agreement that global temperatures are increasing *and* the level of greenhouse gases in the atmosphere is also increasing, the difficulty is to prove a causal link between the two. Most scientists do believe that anthropogenic greenhouse gases are causing global warming and climate change, but there is no universal agreement. Other reasons put forward for the increase in global temperatures include increased solar activity and the effect of volcanic eruptions.

The Intergovernmental Panel on Climate Change (IPCC) stated in its 2007 report: ‘Most of the observed increase in global average temperatures since the mid-20th century is *very likely* due to the observed increase in anthropogenic greenhouse gas concentrations. It is *likely* that there has been significant anthropogenic warming over the past 50 years averaged over each continent (except Antarctica)’.

Evidence of a causal link between increases in the concentrations of greenhouse gases in the atmosphere and global warming can be obtained from complex computer models. These models assume a link between the two variables and are tested against data from the past – if there is good agreement between what is predicted by the models and the trends actually observed, they can be used to make predictions about the effect of future changes on the climate. If there is *then* good agreement between the predictions made by these models (which assume a link between increases in global temperatures and increased levels of greenhouse gases) and future trends, this would strengthen the evidence for a causal link between the two factors.

The effects of climate change

It is difficult to predict the effects of climate change on our planet but some suggested consequences are given here.

- As the Earth’s temperature rises, oceanic water expands – increased sea levels could submerge low-lying areas and many islands. Large populations live in some of these areas. Since 1993, sea levels have been rising at an average rate of more than 3 mm y^{-1} . Only estimates of future rises can be made based on complex models but these predict sea-level rises of up to about 0.5 m during the next 100 years.
- Polar ice caps could melt (Figure C.29) (the melting of floating ice (North Pole) does not cause sea levels to rise).
- Glaciers and snow/ice cover on land could melt – this does increase sea levels.
- The occurrence of extreme weather events such as floods, droughts and heatwaves could increase.
- The amount and distribution of precipitation (rain and snow) could change.



The IPCC is an intergovernmental body that works under the auspices of the United Nations and has 195 countries as members. It reviews and assesses scientific data about climate change in order to provide governments and people around the world with a clearer understanding of the current state of research on climate change and its potential impact. The IPCC does not itself carry out scientific research but thousands of scientists from around the world contribute to it on a voluntary basis.



Figure C.29 Melting sea ice could cause problems for wildlife.

- A warming climate may mean that commercial crops can no longer be produced where they grow now. This could be a massive problem in grain-producing areas that currently produce a large amount of food.
- The distribution of pests and disease-carrying insects could change – e.g. changes in the distribution of the mosquito population could alter the regions where malaria is a danger.

Control of carbon dioxide emissions

The biggest sources of carbon dioxide production in the US are electricity generation, transportation and industry. There are two main approaches to controlling carbon dioxide emissions – reducing the amount of carbon dioxide produced, or capturing and storing the carbon dioxide produced so it is not released to the atmosphere. The effect of deforestation on carbon dioxide levels in the atmosphere should also be considered.

Reducing the amount of carbon dioxide produced

Again there are two approaches – use ways of generating electricity and heat that release less carbon dioxide into the atmosphere, or use less electricity/heat/fossil fuels.

The amount of carbon dioxide produced when electricity is generated can be reduced by switching from using coal-fired power stations to natural gas or, especially, nuclear power. Coal produces the most carbon dioxide per kWh of electricity generated. Also, increased amounts of renewable energy sources such as hydroelectric power, solar energy and wind power could be used.

Wood-burning power stations could reduce emissions because burning the wood simply replaces the carbon dioxide that was removed when the trees grew. How close to carbon-neutral this is depends on various factors such as how far the wood must be transported etc. before being burned. Trees must also be replanted after they are harvested because deforestation reduces the amount of carbon dioxide removed from the atmosphere.

The amount of heat/electricity/fossil fuels we need can be reduced in many ways, for example:

- Design more fuel-efficient buildings by using more insulation and/or active and passive solar heating.
- Turn off lights when not needed and do not leave electrical devices on standby.
- Design electrical devices that use less electricity such as energy-efficient light bulbs and more energy-efficient refrigerators.
- Develop vehicles that are more fuel-efficient. The European Union has introduced legislation that attempts to reduce the average carbon dioxide emissions from all new vehicles produced from 2015 to less than 130 g km^{-1} . This should then be reduced to 95 g km^{-1} by 2020.
- Develop the use of hybrid, fuel cell and electric vehicles. However, the carbon footprint of electric vehicles and those using hydrogen fuel cells depends on how the electricity used to power them is generated and how the hydrogen is produced.
- Encourage the use of public transport, cycling etc.



The effect of global warming on global food production is difficult to estimate but, depending on the extent of the temperature change, it could actually result in an overall increase in the potential for global food production.



The Kyoto Protocol is an international treaty signed in 1997 which attempted to set binding targets for reduction of carbon dioxide emissions by developed countries. The protocol was accepted and ratified by the governments in many countries but some, such as the US and Canada, subsequently withdrew. The Kyoto Protocol was due to expire in 2012 but at a conference in Doha, Qatar it was extended to 2020. However, further withdrawal of countries and the non-inclusion of China means that the treaty only covers countries responsible for less than 20% of the carbon dioxide produced globally.

Carbon capture and storage

Various schemes have been proposed to capture carbon dioxide from the flue gases of power stations, before it can be released into the atmosphere, and storing it in some way – this is called carbon dioxide capture and sequestration, or carbon capture and storage (CCS). Some processes that are already in operation involve capturing the carbon dioxide from power station emissions, transporting it through pipelines and storing it underground, where it is trapped. The carbon dioxide can also be used in enhanced oil recovery which involves pumping carbon dioxide underground to force more oil out of an oil well (carbon dioxide can also mix with the oil to reduce its viscosity). Some of the carbon dioxide does return to the surface with the oil but this is reused and not released to the atmosphere.

Other proposed schemes for CCS include reacting carbon dioxide with minerals underground to form insoluble carbonates. The downside to all these processes is that they will require energy and cost money. However, if there is a net decrease in the amount of carbon dioxide released to the atmosphere it could be worth it.

Deforestation

Deforestation refers to the action of cutting down trees to clear areas of land (Figure C.30), usually so that they can be used for agricultural purposes. Deforestation is a major contributor to carbon dioxide emissions and it has been estimated that this could be responsible for up to 17% of global anthropogenic carbon dioxide emissions. Trees remove carbon dioxide from the atmosphere as they grow, but when they are cut down there are two effects – not only does this removal of carbon dioxide stop, but also the carbon dioxide that they removed is returned to the atmosphere when they are burned or left to rot.



Figure C.30 Deforestation in Indonesia.

Global dimming

Global dimming refers to the decrease in the amount of electromagnetic radiation from the Sun that reaches the surface of the Earth due to the presence of particulates in the atmosphere. These small particles reflect sunlight back into space and therefore cause cooling of the Earth.

Particulates are solid or liquid particles in the air – the term for such particles suspended in a gas is ‘aerosol’. Anthropogenic sources of particulate matter include:

- smoke/soot from incomplete combustion of wood, coal, petrol and diesel
- fly ash – contains soot and metal oxides from fossil fuels burning in furnaces
- dust from mechanical activity, demolition, metal-working etc.

Natural sources of particulate matter include pollen, dust, smoke/soot from forest fires and sea spray. The eruption of volcanoes results in large amounts of dust and sulfur dioxide entering the atmosphere. The sulfur dioxide causes the formation of tiny droplets of sulfuric acid in the upper atmosphere – a sulfate aerosol, which is very good at reflecting sunlight.

Not only can particulates reflect sunlight back into space but they can also have an indirect effect in that they can change the make-up of clouds. Water droplets form around particulates and this reduces the size of water droplets as clouds form. Clouds also cause global dimming by reflecting

sunlight back into space but the polluted clouds, with smaller water droplets, reflect more sunlight.

Ocean acidification

Increasing levels of carbon dioxide in the atmosphere resulting from human activity are making the oceans more acidic. Carbon dioxide is slightly soluble in water and an equilibrium is set up between carbon dioxide in the atmosphere and aqueous carbon dioxide in the oceans:

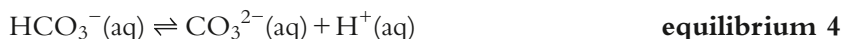
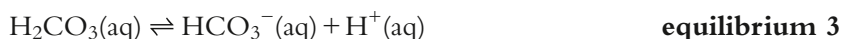


This is called a heterogeneous equilibrium because it involves substances in two different physical states. As the amount of carbon dioxide in the atmosphere increases, this will shift the position of the above equilibrium to the right – therefore there will be more carbon dioxide dissolved in the oceans.

The dissolved CO_2 reacts with water to form carbonic acid:



Carbonic acid is a weak diprotic acid and can dissociate:



As more carbon dioxide dissolves, the position of equilibrium 2 shifts to the right, which increases the concentration of H_2CO_3 ; this causes the position of equilibrium 3 to shift to the right and therefore increases the concentration of the $\text{H}^+(\text{aq})$ ions and hence the acidity of the water.

‘Ocean acidification’ is the reduction of the pH of the oceans over an extended period of time through the increased uptake of carbon dioxide from the atmosphere. Some scientists believe that the pH of the oceans could fall by up to 0.4 units by the end of this century.

There are many possible consequences associated with ocean acidification such as possible effects on the ability of corals, shellfish, phytoplankton etc. to make calcium carbonate shells and skeletons, a decrease in the metabolic rate and activity of jumbo squid and so on.

Nature of science

The ideas of cause and effect are very important in science. There is a definite correlation between the level of carbon dioxide in the atmosphere and global temperature but it is very difficult to prove that increases in anthropogenic carbon dioxide cause increases in global temperature. There have, for instance, been wide variations in the Earth’s temperature in the fairly recent past, for example the ‘Medieval warm period and the ‘little ice age’. Scientists understand the mechanism by which an increased level of carbon dioxide could cause global warming but this still does not prove a link because planet Earth is an extremely complex system and there are many variables that must be considered. Computer modelling, which uses complex mathematical equations and many approximations, assumptions and simplifications, is an important research tool in the search for more evidence.



The issues concerning man-made climate change are very complex – how do we make decisions about this? To what extent are our decisions based on our trust in the knowledge and understanding of experts in the field, our own understanding or emotion? Is there a tendency to just align ourselves with an organisation (or individual) which has similar views to our own and then oppose all other views? To what extent are our opinions objective? How much information do we need to make a decision – do we need to examine all the primary data? What if we are not capable of understanding the primary data or it is not available?

? Test yourself

- 17 Give the names of **three** greenhouse gases produced by human activity.
- 18 State whether each of the following statements is true or false:
- a Methane causes global dimming.
 - b Increased levels of carbon dioxide in the atmosphere cause the pH of the oceans to increase.
 - c More carbon dioxide than methane is produced each year by human activities.
 - d Deforestation probably contributes to climate change.

C6 Electrochemistry, rechargeable batteries and fuel cells (HL)

C6.1 Batteries and fuel cells

Batteries

A battery is a portable electrochemical device that produces electricity – it is made up of one or more voltaic cells connected in series.

The term ‘electrochemical cell’ can refer to either voltaic (Galvanic) cells or electrolytic cells.

Voltaic cells

We have already met voltaic cells in Topic 9. Figure C.31 shows a simple voltaic cell – in general, a cell consists of two different electrodes dipping into an electrolyte.

Cells involve chemical reactions (redox reactions) in which the negative electrode releases electrons as it reacts (is oxidised) while the positive electrode gains electrons (is reduced). A cell ‘runs out’ when at least one of the chemicals involved in these reactions has been used up.

The voltage of a cell depends primarily on the materials used to make the cell.

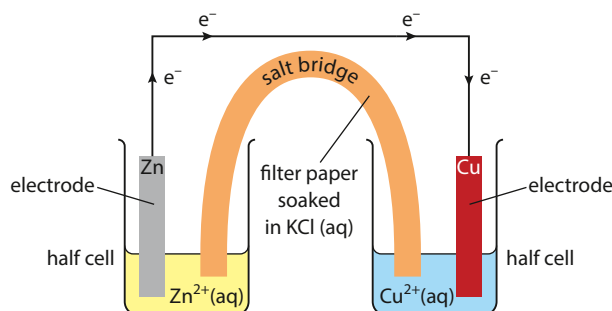


Figure C.31 A typical voltaic cell.

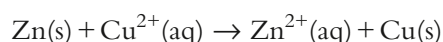
Learning objectives

- Understand that the voltage of a voltaic cell depends on the materials from which it is made
- Understand that the total amount of electrical energy (work) that can be obtained from a cell depends on the quantities of substances used in the cell
- Understand the internal resistance of a voltaic cell in terms of the diffusion of ions
- Distinguish between primary cells, secondary cells and fuel cells
- Understand how fuel cells work
- Write equations for the reactions at the electrodes when fuel cells operate with acidic and alkaline electrolytes
- Understand that microbial fuel cells (MFCs) represent a possible sustainable energy source using organic matter in waste water
- Calculate the thermodynamic efficiency of fuel cells
- Understand the reactions at the anode and cathode in rechargeable batteries
- Discuss the advantages of different types of cells in terms of size, mass and voltage
- Compare fuel cells with rechargeable batteries

If the zinc in the cell in Figure C.31 is replaced by magnesium, the voltage of the cell will increase because magnesium has a more negative standard electrode potential than zinc. The nature and concentration of the electrolyte can also affect the voltage.

The total amount of electrical energy (work) that can be obtained from a cell depends on the quantities of substances used in the cell.

In the cell in Figure C.31 the overall reaction that occurs is:



Zinc metal and copper ions are used up and so if more of these are used initially a greater total amount of energy will be available from the cell.

All voltaic cells have an internal resistance which limits the maximum current that can be obtained from the cell. Current is the rate of flow of charge and the internal resistance arises because of the finite amount of time that it takes for ions to diffuse to the electrodes in the cell. In the cell in Figure C.32 electrons are transferred to Cu^{2+} ions at the copper electrode. Once copper ions at the electrodes have been reduced to copper, more copper ions must diffuse in to replace them – this takes a certain amount of time and therefore limits the number of electrons that can be removed from the electrode per second, and hence the current.

Different types of cells

Primary cells cannot usually be recharged using mains electricity – the reaction in the cell is non-reversible.

Primary cells are a very expensive source of electrical energy compared to mains electricity but are used because of their convenience and portability. Examples of primary cells include zinc–carbon cells/batteries, often used in torches etc., and alkaline cells/batteries.

A secondary cell is one that can be recharged using mains electricity and is often called a rechargeable battery. The chemical reactions in a rechargeable battery are reversible and can be reversed by connecting them to an electricity supply.

Secondary cells include the lead–acid battery used in car engines and lithium-ion batteries used in laptops etc.

A fuel cell differs from other cells in that it uses a continuous supply of reactants from an external source.

Fuel cells use the reaction between a fuel (such as hydrogen or methanol) and an oxidising agent (e.g. oxygen) to produce electrical energy directly. Hydrogen fuel cells are used in, for instance, hydrogen-powered buses (Figure C.33) and were used in the space shuttle.

All cells involve the conversion of chemical energy to electrical energy.

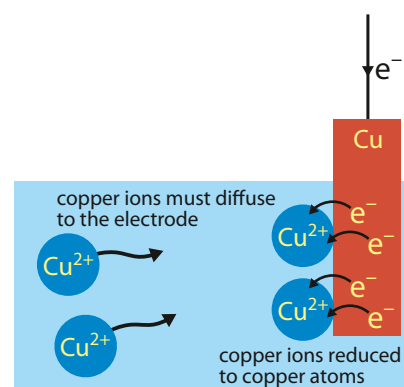


Figure C.32 The time it takes for diffusion of ions in the electrolyte causes internal resistance.



Figure C.33 A London bus powered by hydrogen fuel cells.

Fuel cells

A fuel cell is a specialised type of electrochemical cell in which a fuel and an oxidising agent react in the presence of an electrolyte to produce electrical energy. In the hydrogen–oxygen fuel cell, hydrogen reacts with oxygen to produce water and electricity.

Figure C.34 shows an alkaline hydrogen–oxygen fuel cell with a potassium hydroxide electrolyte. Electrodes are typically made of carbon and incorporate a metal catalyst such as nickel.

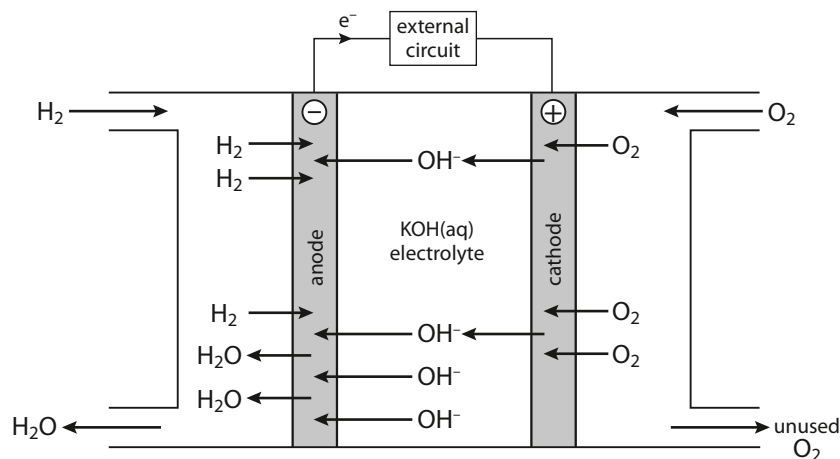
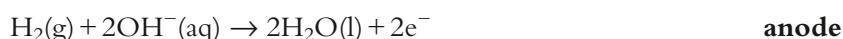


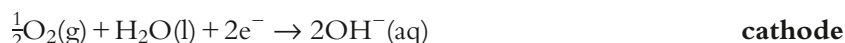
Figure C.34 An alkaline hydrogen–oxygen fuel cell with a potassium hydroxide electrolyte.

The two electrodes are separated by a porous matrix saturated with an electrolyte which may be either an alkaline or an acidic solution. In the case of an alkaline electrolyte, typically a solution of potassium hydroxide (KOH) is used.

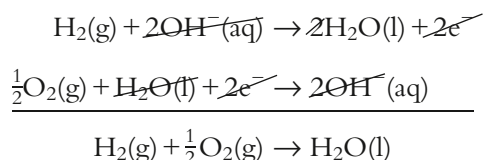
At the anode, hydrogen undergoes oxidation when it reacts with hydroxide ions (OH^-) to form water:



Electrons are released in this oxidation process and flow around the external circuit to the cathode. At the cathode, oxygen undergoes reduction when it reacts with water and gains electrons from the electrode:



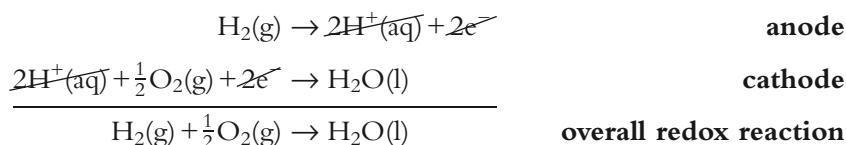
If these reactions are added together and common terms cancelled we get:



overall redox reaction

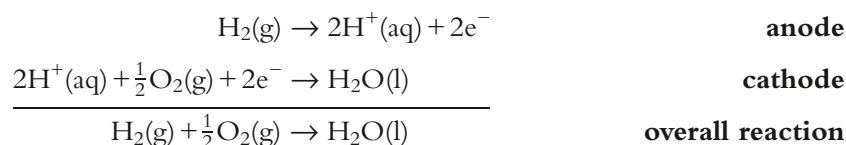
A fuel cell involves a reaction equivalent to a combustion reaction, but releases electrical energy rather than heat energy. The separation of the oxidation and reduction reactions permits the flow of electrons and the generation of an electric current.

In an acidic fuel cell the alkaline electrolyte is replaced by an acidic solution, commonly phosphoric acid. The reactions occurring at the anode and cathode in an acidic fuel cell are different:



Proton exchange membrane (PEM) fuel cells

This type of fuel cell is the one that is increasingly finding use in hydrogen-powered vehicles. Hydrogen passes through a porous carbon anode (Figure C.35) that contains a platinum catalyst and is oxidised to form hydrogen ions (protons) and electrons (which travel through the external circuit). The hydrogen ions travel through the PEM to the cathode (containing a platinum catalyst) where they combine with oxygen molecules and electrons to form water. The reactions that occur are the same as in the phosphoric acid fuel cell:



The fuel cells described above rely on external sources of a fuel and an oxidising agent, which are depleted during use and require constant replenishing. They are therefore very different from primary cells that contain stored chemical energy within them. Once a primary cell runs out it must be disposed of, but a hydrogen fuel cell will never run out as long as more hydrogen and oxygen are supplied.

In a hydrogen–oxygen fuel cell in a vehicle, the hydrogen must be stored on board in some way and the oxygen may come from purified air or an oxygen tank. Hydrogen–oxygen fuel cells do not produce any pollution because the only products are water, electricity and a small amount of heat. However, the production of hydrogen from the electrolysis of water, for example, must be considered when assessing the overall environmental impact of fuel-cell use.

A proton exchange membrane (PEM) is sometimes called a polymer electrolyte membrane.

Because of the use of platinum catalysts and precision engineering, hydrogen–oxygen fuel cells are currently very expensive to make.

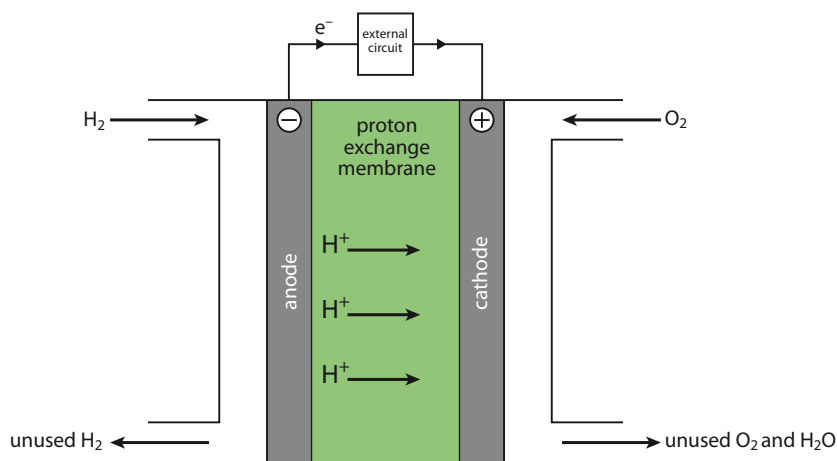
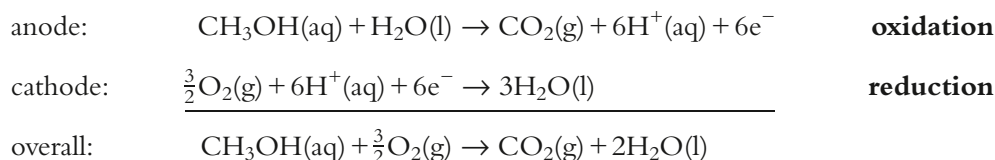


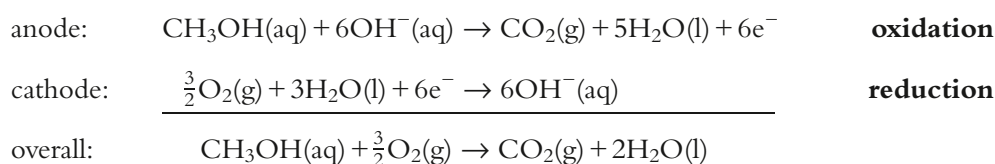
Figure C.35 A PEM fuel cell.

Methanol fuel cells

One disadvantage of using hydrogen fuels cells is the difficulty of storing hydrogen fuel (a gas) and one solution to this problem is to use a liquid fuel, such as methanol. A direct methanol fuel cell works in a similar way to the PEM fuel cell described above. The anode contains a platinum–ruthenium catalyst, the catalyst at the cathode is platinum and there is a PEM between the electrodes. Methanol solution and oxygen are fed into the fuel cell, and the equations for the reactions are:



A methanol fuel cell can also be constructed using an alkaline electrolyte and the reactions are:



Again, the overall reaction is equivalent to a combustion reaction.

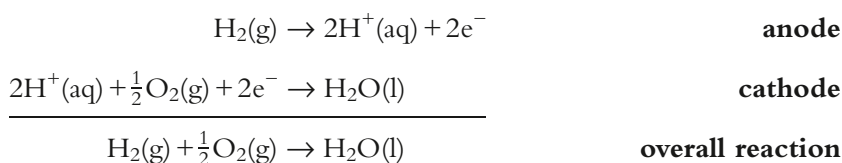
Direct methanol fuel cells are being developed for use in laptop computers and mobile phones. The fuel cell is built into the device and a methanol cartridge is plugged in to provide the fuel. When the fuel has been used up the cartridge is disposed of and replaced by a new one. The use of this sort of technology allows laptops and phones to be used well away from the electricity grid and has many potential military and civilian applications.

Reformed methanol fuel cells

In this type of fuel cell, the methanol is first reformed to hydrogen and carbon dioxide:



The hydrogen is then used as the fuel for the fuel cell. The reactions that occur are the same as in the PEM fuel cell:



Microbial fuel cells

In an MFC, bacteria attach themselves to the anode and when they oxidise organic matter present in, say, waste water, they produce carbon dioxide, protons (H^+ ions) and electrons – the electrons produced are transferred from the bacteria to the electrode. The reactions in the anode chamber (Figure C.36) occur in the absence of oxygen (anaerobic conditions).

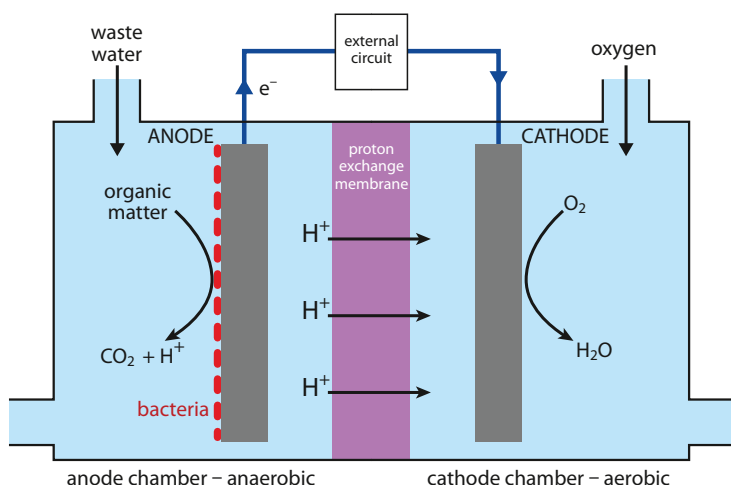
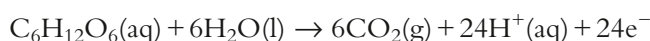
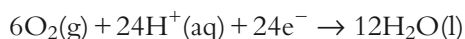


Figure C.36 A microbial fuel cell.

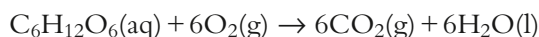
For instance, the oxidation of a carbohydrate such as glucose can be represented as:



The electrons travel through the external circuit to the cathode, while the protons pass through a PEM to the cathode compartment, where they combine with oxygen and electrons to form water.



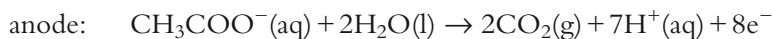
The overall reaction is:



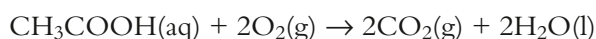
which is equivalent to the combustion of glucose.

The microbial fuel cell represents a possible sustainable source of electrical energy, produced from the oxidation of carbohydrates and other substrates in waste water. They provide two functions – they oxidise the organic matter in waste water, which is one of the important stages in the treatment of waste water, and they generate electricity at the same time.

Geobacter bacteria such as *Geobacter sulfurreducens* are used in some fuel cells to reduce ethanoate ions, CH_3COO^- , under anaerobic conditions. The reactions occurring in the anode and cathode chambers can be represented by:



The actual reactions occurring are substantially more complicated than this. The overall reaction is equivalent to the complete combustion of ethanoic acid:



Thermodynamic efficiency of fuel cells

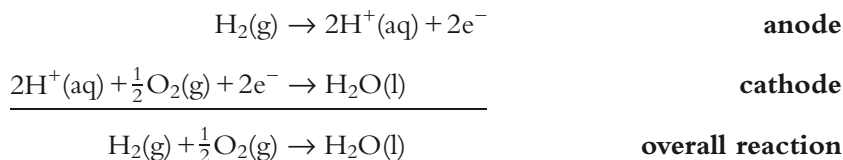
The thermodynamic efficiency of a fuel cell can be calculated using the free energy change for the cell reaction and the enthalpy change for the equivalent combustion reaction.

$$\text{Thermodynamic efficiency} = \frac{\Delta G}{\Delta H}$$

The change in free energy represents the maximum amount of work that can be obtained from a cell and, as shown in Topic 9, can be calculated from:

$$\Delta G^{\ominus} = -nFE^{\ominus}$$

Consider the PEM fuel cell reactions:



The maximum voltage for this cell is 1.23 V and two electrons are transferred, so ΔG^{\ominus} can be worked out from:

$$\begin{aligned}\Delta G^{\ominus} &= -nFE^{\ominus} \\ &= -2 \times 96\,500 \times 1.23 \\ &= -237\,000 \text{ J mol}^{-1} \text{ or } -237 \text{ kJ mol}^{-1}\end{aligned}$$

The amount of heat energy given out in the combustion of one mole of hydrogen is simply the standard enthalpy change of combustion of hydrogen (or the standard enthalpy change of formation of water), which is -286 kJ mol^{-1} .

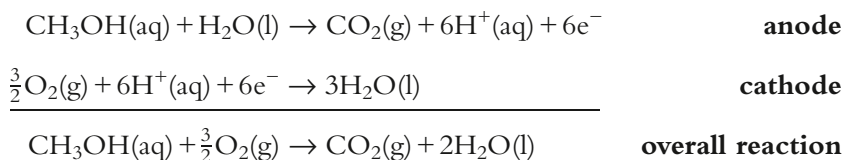
The thermodynamic efficiency is calculated from:

$$\frac{\Delta G}{\Delta H} = \frac{237}{286} = 0.829$$

This can be multiplied by 100 to give a percentage efficiency of 82.9%.

This is the maximum possible efficiency of the fuel cell under these conditions. It is not 100% because of entropy – some energy is unavailable to do work because the reaction involves a decrease in entropy. Energy is needed to cause the redistribution of the available energy among the various energy states of the products – the energy is less spread out or more concentrated in the products.

We can carry out a similar calculation for the direct methanol fuel cell.



The maximum voltage is 1.21 V and six electrons are transferred so:

$$\begin{aligned}\Delta G^{\ominus} &= -nFE^{\ominus} \\ &= -6 \times 96\,500 \times 1.21 \\ &= -701\,000 \text{ J mol}^{-1} \text{ or } -701 \text{ kJ mol}^{-1}\end{aligned}$$

The enthalpy change of combustion of methanol is -726 kJ mol^{-1} and:

$$\frac{\Delta G}{\Delta H} = \frac{701}{726} = 0.965 \text{ or } 96.5\%$$

The value for the enthalpy change of combustion of liquid methanol has been used here.

This value is higher than the previous one because the reaction involves a smaller decrease in entropy (smaller change in the number of moles of gas). If a fuel cell reaction involves an increase in entropy, the theoretical efficiency of the fuel cell will be greater than 1. The values here represent the maximum possible efficiency and operating efficiency is significantly lower than these values.

Rechargeable batteries

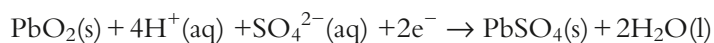
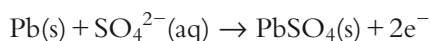
Rechargeable batteries, or secondary cells, are a portable type of electrochemical cell that generates a current via electrically reversible reactions.

Lead–acid battery

One of the oldest forms of rechargeable cell is the lead–acid battery, still used in the majority of car engines. The lead–acid battery consists of lead anodes and cathodes made from lead covered with a layer of lead(IV) oxide (PbO_2) in a concentrated sulfuric acid electrolyte.

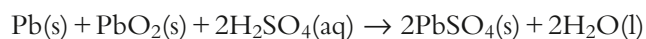
A typical car battery is made up of six electrical cells in series and provides a potential difference of about 12 V. As the battery discharges, lead at the anode is oxidised, releasing electrons, while the lead(IV) oxide at the cathode is reduced. The product of both reactions is solid lead(II) sulfate, which accumulates on the electrodes within the battery.

During discharge:

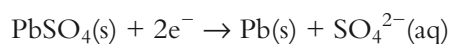


overall reaction

This can also be written as:



When a battery loses its charge, the lead(II) sulfate can be converted back to lead and lead(IV) oxide by applying an external electrical source – the chemical reactions above are reversed. For example, the reaction that occurs at the electrode attached to the negative side of the charging supply (cathode) is:



Lead–acid storage batteries are quite large and heavy and are unsuitable for uses where they have to be carried around.

Nickel–cadmium battery

NiCad batteries are often seen as the classic small rechargeable ones (sizes AAA to D) used for powering small electronic devices such as clocks, calculators, remote controls and toys. They typically generate smaller voltages than lead–acid batteries – around 1.2 V. NiCad batteries use a nickel oxide hydroxide [NiO(OH)] cathode and a metallic cadmium anode separated by a potassium hydroxide electrolyte.

Lead–acid cells are also used as emergency sources of power (standby power). This is vital in places like hospitals – the emergency power cuts in if the mains electricity supply fails. They are also used in theatres, cinemas (and schools) to guarantee emergency lighting at all times.

Exam tip

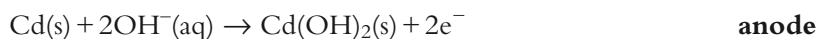
When the battery is discharging the anode is the negative electrode and the cathode the positive electrode.

When a car is in motion, the battery is being charged because it is linked to the car's alternator.

Exam tip

When charging the battery the anode and cathode labels must be changed around – the cathode is the electrode at which reduction occurs. The cathode is thus the electrode attached to the negative side of the electricity supply.

During discharge:



These reactions are reversed when the battery is recharged.

Lithium-ion battery

These batteries are commonly used as rechargeable cells in mobile telephones, laptop computers and high-energy usage portable electronic devices. They can generate higher voltages (around 3–4V) than NiCad batteries. Typically, the anode is made of carbon with lithium atoms inserted in the lattice, and the cathode is a metal oxide such as manganese(IV) oxide or cobalt(IV) oxide with Li^+ ions inserted in the lattice. The electrolyte is a complex lithium salt such as lithium hexafluorophosphate (LiPF_6) dissolved in an organic solvent (Figure C.37).

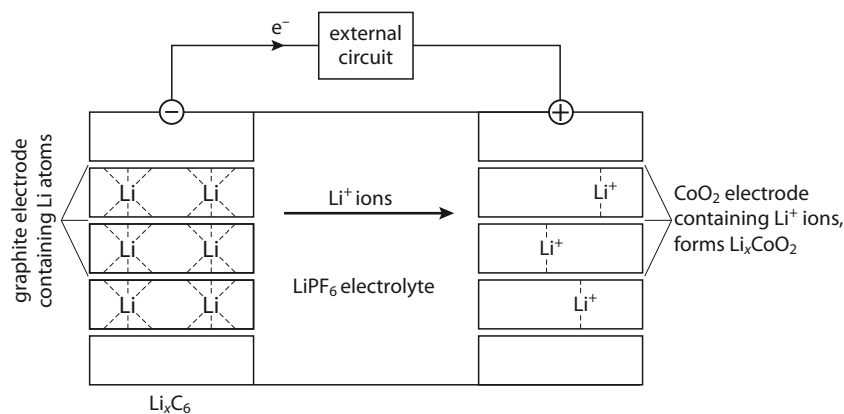


Figure C.37 A lithium-ion battery – lithium ions are exchanged between complexed lithium and graphite electrodes.

During discharge, lithium atoms are oxidised at the negative electrode and electrons are released to the external circuit:



where the equation represents molar quantities reacting and x is a number less than 1.

The lithium ions move through the electrolyte to the cathode, where they become inserted into the lattice of the transition metal oxide:



The overall redox reaction during discharge is:



The net effect of these reactions is that lithium atoms are oxidised at the negative electrode and cobalt ions are reduced at the positive electrode. For every lithium ion present in Li_xCoO_2 , the oxidation number of a cobalt ion must be reduced by 1 from +4 to +3.

The reverse reactions occur during charging, and lithium ions move in the opposite direction.

An Li^+ ion and a Co^{3+} ion together give a charge of 4+ to cancel out the charge on two O^{2-} ions. Therefore if the formula of the species formed at the cathode is $\text{Li}_{0.2}\text{CoO}_2$ what we have present is $(\text{Li}^+)_{0.2}(\text{Co}^{4+})_{0.8}(\text{Co}^{3+})_{0.2}(\text{O}^{2-})_2$.

Both nickel–cadmium and lithium–ion batteries are much smaller and lighter than lead–acid storage batteries and are, therefore, suitable for use in portable devices.

Similarities and differences between fuel cells and rechargeable batteries

Similarities:

- They both generate electrical energy from chemical energy.
- They are both composed of an anode, a cathode and an electrolyte.
- They both generate a current based on the separation of reduction and oxidation reactions and the flow of electrons through an external circuit.

Differences:

- Fuel cells require an external source of chemical energy (fuel), but rechargeable batteries have their chemical energy source within them.
- Fuel cells never run out so long as there is a constant supply of fuel from an external source; rechargeable batteries run out and then have to be recharged by connecting to an electricity supply which reverses the reactions.
- Rechargeable batteries are far cheaper than fuel cells.
- Fuel cells are capable of generating a far greater quantity of electricity than rechargeable batteries because the fuel is supplied constantly.
- Fuel cells are non-polluting – hydrogen fuel cells produce water as the only product; rechargeable batteries may contain toxic metals (Cd, Pb) so are difficult to dispose of.
- Fuel cells can produce drinkable water as a byproduct; there is no byproduct of rechargeable batteries.
- Rechargeable batteries can only be recharged so many times – they have a finite life and must be replaced eventually; fuel cells have much better longevity.

Comparison of different types of cell

Table C.11 compares the various types of cell discussed above in terms of size, mass and voltage of the cells.

Cell type	Size/mass	Voltage
Fuel cells	Can be made in a variety of sizes. Fuel cell stacks used to power buses are quite large and heavy, but small fuel cells have been developed for laptops etc. Some fuel cells have a very high power-to-mass ratio and high power-to-volume ratio.	Typically about 0.6–0.8 V but depends on the type of fuel cell. Used in a stack with multiple fuel cells connected in series.
Lead–acid batteries	Large and heavy – not suitable for portable devices. Low power-to-mass ratio and power-to-volume ratio.	2 V per cell – usually used as a battery with six cells in series
Nickel–cadmium batteries	Small and light – used in a variety of portable devices. Intermediate power-to-mass ratio and power-to-volume ratio.	1.2 V
Lithium-ion batteries	Small and light – used in a variety of portable devices. Produce the greatest amount of power per unit mass or unit volume of the rechargeable batteries.	About 3.7 V – the highest voltage per cell.

Table C.11 A comparison of different types of cell.



Figure C.38 This car is powered by rechargeable batteries.

Much research is being carried out into developing electric vehicles (Figure C.38). These have advantages over gasoline-powered vehicles in that they do not produce carbon dioxide when they are operating – although carbon dioxide might be produced when the charging electricity is generated or when producing hydrogen for a fuel cell.

The two main possibilities for powering vehicles by electricity are to use rechargeable batteries (such as the lithium-ion battery) that can be plugged in and recharged every so often, or to use fuel cells and a hydrogen storage tank which will have to be refilled every so often.

When considering which is the better solution, the two important factors that must be looked at are the mass of the battery/fuel cell and the volume of space it occupies. For fuel cells, the mass/volume occupied by the fuel tank and fuel must also be factored into the calculation – hydrogen must be stored under pressure to reduce its volume and the fuel tank must be reinforced (and so can be quite heavy) to withstand high pressures. There is also safety to consider in the use of hydrogen and other methods of storing hydrogen are being researched.

The energy-to-mass ratio (specific energy) of a fuel cell depends on the type of fuel cell stack but fuel cells used to power vehicles have a higher energy-to-mass ratio than rechargeable batteries which means that the vehicles should be able to travel further for the same mass of a battery/fuel cell + fuel. The advantage of using fuel cells to power vehicles increases with the distance between stops that is required because this can be increased by simply having a larger fuel tank which will not add a large amount of mass (hydrogen is a very light gas). With rechargeable batteries the distance between stops can only be improved by adding extra rechargeable batteries which contribute significantly to the overall mass of the vehicle and take up a lot of space.

The energy density (energy per unit volume) of fuel cells, including the hydrogen tank, is similar to that of lithium-ion batteries – but again, the longer the distance required between stops for refuelling, the greater the advantage of using fuel cells.

Test yourself

- 19 Write ionic half-equations for the following reactions:
 - a the reaction that occurs at the electrode attached to the negative side of a power supply when a lead–acid battery is charged
 - b the reaction that occurs at the electrode attached to the positive side of a power supply when a NiCad battery is charged.
- 20 Write an overall equation for the reaction that occurs when a lithium-ion cell is charged.

C6.2 The effect of concentration on electrode potentials

The Nernst equation

In Topic 9 we used standard electrode potentials to calculate standard cell potentials. Standard conditions for these measurements are taken as solutions of concentration 1 mol dm^{-3} and gases at a pressure of 100 kPa (1 bar). The **Nernst equation** can be used to calculate the electrode potentials of half-cells or cells under non-standard conditions:

$$E = E^\ominus - \left(\frac{RT}{nF} \right) \ln Q$$

where

E is the electrode potential

E^\ominus is the standard electrode potential

R is the gas constant, $8.31 \text{ J K}^{-1} \text{ mol}^{-1}$

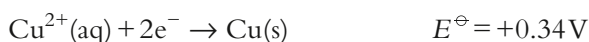
T is the temperature in kelvin

n is the number of electrons transferred

F is the Faraday constant, $96\,500 \text{ C mol}^{-1}$

Q is the reaction quotient (introduced in Topic 7)

Consider the reduction of copper(II) ions:



We can work out the electrode potential when the concentration of copper ions is 0.10 mol dm^{-3} and the temperature is 298 K as follows:

$$\text{The reaction quotient is } Q = \frac{[\text{Cu}(\text{s})]}{[\text{Cu}^{2+}(\text{aq})]}$$

In a heterogeneous reaction, the concentration of a pure solid (or pure liquid) is taken as 1 so we have:

$$\begin{aligned} Q &= \frac{1}{[\text{Cu}^{2+}(\text{aq})]} \\ &= \frac{1}{0.10} \\ &= 10 \end{aligned}$$

The number of electrons transferred is 2 and so we have:

$$\begin{aligned} E &= E^\ominus - \left(\frac{RT}{nF} \right) \ln Q \\ &= +0.34 - \left(\frac{8.31 \times 298}{2 \times 96\,500} \right) \ln 10 \\ &= +0.31 \text{ V} \end{aligned}$$

It can be seen from this that changing the concentration does not have much effect on an electrode potential, unless the change in concentration is very large.

There are two approaches to working out a cell potential under non-standard conditions and this is best illustrated with a worked example.

Learning objectives

- Understand how the Nernst equation can be used to calculate electrode potentials under non-standard conditions
- Understand how concentration cells work
- Solve problems involving the Nernst equation

Exam tip

A temperature is not actually included in the definition of standard conditions but where one is not specified it can be taken as 298 K .

Exam tip

You can use ideas similar to Le Chatelier's principle to work out whether your answer is sensible or not. If we regard $\text{Cu}^{2+}(\text{aq}) + 2\text{e}^- \rightleftharpoons \text{Cu}(\text{s})$ as an equilibrium system, decreasing the concentration of $\text{Cu}^{2+}(\text{aq})$ will shift the position of the equilibrium to the left, and so the forward reaction becomes less favourable and the value of the electrode potential becomes less positive.

Worked examples

C.7 Calculate the cell potential of the $\text{Zn(s)}|\text{Zn}^{2+}(\text{aq})||\text{Ag}^{+}(\text{aq})|\text{Ag(s)}$ cell at 25°C when the concentration of $\text{Zn}^{2+}(\text{aq})$ is 0.200 mol dm^{-3} and that of $\text{Ag}^{+}(\text{aq})$ is 0.100 mol dm^{-3} .

The first approach to this is to work out each individual cell potential under the non-standard conditions and then combine them.

The standard electrode potentials are:



Ag^{+}/Ag half-cell

The reaction quotient is $Q = \frac{[\text{Ag(s)}]}{[\text{Ag}^{+}(\text{aq})]}$

In a heterogeneous reaction the concentration of a solid is taken as 1 so we have:

$$\begin{aligned} Q &= \frac{1}{[\text{Ag}^{+}(\text{aq})]} \\ &= \frac{1}{0.100} \\ &= 10.0 \end{aligned}$$

The number of electrons transferred is 1 and so we have:

$$\begin{aligned} E &= E^{\ominus} - \left(\frac{RT}{nF} \right) \ln Q \\ E &= 0.80 - \left(\frac{8.31 \times 298}{1 \times 96\,500} \right) \ln 10.0 \\ &= 0.74\text{ V} \end{aligned}$$

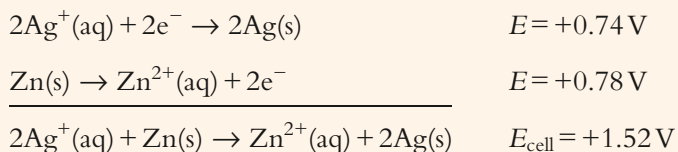
Zn^{2+}/Zn half-cell

$$\begin{aligned} Q &= \frac{1}{[\text{Zn}^{2+}(\text{aq})]} \\ &= \frac{1}{0.200} \\ &= 5.00 \end{aligned}$$

The number of electrons transferred is 2 and so we have:

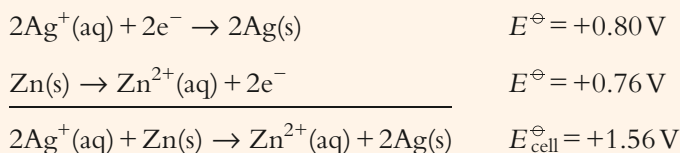
$$\begin{aligned} E &= E^{\ominus} - \left(\frac{RT}{nF} \right) \ln Q \\ &= -0.76 - \left(\frac{8.31 \times 298}{2 \times 96\,500} \right) \ln 5.00 \\ &= -0.78\text{ V} \end{aligned}$$

In order to combine the half equations to give the overall cell reaction the $\text{Zn}^{2+}|\text{Zn}$ half-cell is reversed and the $\text{Ag}^+|\text{Ag}$ half-cell reaction is multiplied by 2 to balance out the electrons:



Remember, from Topic 9, that cell potentials are not multiplied by anything when they are combined to give the cell potential.

An alternative approach is to work out the standard cell potential first and then use the Nernst equation to work out the cell potential under non-standard conditions.



The reaction quotient is worked out in the same way as an equilibrium constant and so the concentrations are raised to the powers of the coefficients in the equation. Concentrations of solids are taken as 1.

$$\begin{aligned} Q &= \frac{[\text{Zn}^{2+}(\text{aq})]}{[\text{Ag}^+(\text{aq})]^2} \\ &= \frac{0.200}{0.100^2} \\ &= 20.0 \end{aligned}$$

The number of electrons transferred is 2 and so we have:

$$\begin{aligned} E &= E^\ominus - \left(\frac{RT}{nF} \right) \ln Q \\ &= +1.56 - \left(\frac{8.31 \times 298}{2 \times 96\,500} \right) \ln 20.0 \\ &= 1.52 \text{ V} \end{aligned}$$

C.8 Work out the electrode potential of the $\text{Cr}_2\text{O}_7^{2-}(\text{aq})|\text{Cr}^{3+}(\text{aq})$ half-cell at 298 K given that the concentration of $\text{Cr}_2\text{O}_7^{2-}(\text{aq})$ is $0.0200 \text{ mol dm}^{-3}$, that of Cr^{3+} is $0.0600 \text{ mol dm}^{-3}$ and the pH is 3.00.

The half-cell reaction is $\text{Cr}_2\text{O}_7^{2-}(\text{aq}) + 14\text{H}^+(\text{aq}) + 6\text{e}^- \rightleftharpoons 2\text{Cr}^{3+}(\text{aq}) + 7\text{H}_2\text{O}(\text{l})$ and $E^\ominus = 1.33 \text{ V}$.

A pH of 3.00 corresponds to $[\text{H}^+(\text{aq})] = 10^{-3.00} \text{ mol dm}^{-3}$

$\text{H}_2\text{O}(\text{l})$ is a pure liquid and its concentration is taken as 1 in the expression for Q :

$$\begin{aligned} Q &= \frac{[\text{Cr}^{3+}(\text{aq})]^2}{[\text{Cr}_2\text{O}_7^{2-}(\text{aq})][\text{H}^+(\text{aq})]^{14}} \\ &= \frac{0.0600^2}{0.0200 \times (10^{-3})^{14}} \\ &= 1.80 \times 10^{41} \end{aligned}$$

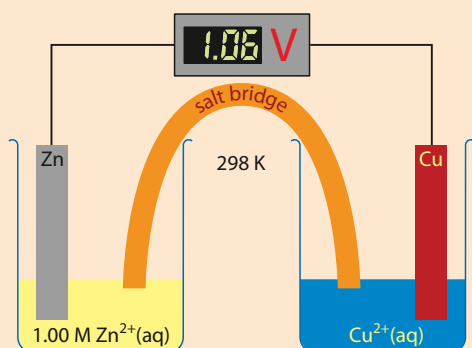
The number of electrons transferred is 6 and so we have:

$$\begin{aligned}
 E &= E^\ominus - \left(\frac{RT}{nF} \right) \ln Q \\
 &= 1.33 - \left(\frac{8.31 \times 298}{6 \times 96\,500} \right) \ln(1.80 \times 10^{41}) \\
 &= 0.924 \text{ V}
 \end{aligned}$$

The Nernst equation can also be used to work out solution concentrations.

Worked example

C.9 A voltaic cell is set up as show in the diagram. The concentration of $\text{Zn}^{2+}(\text{aq})$ is 1.00 mol dm^{-3} but that of $\text{Cu}^{2+}(\text{aq})$ is unknown. The cell potential is 1.06 V – calculate the concentration of the copper(II) solution.



The overall reaction is $\text{Zn}(\text{s}) + \text{Cu}^{2+}(\text{aq}) \rightarrow \text{Zn}^{2+}(\text{aq}) + \text{Cu}(\text{s})$.

The standard cell potential can be worked out from the standard electrode potentials to be 1.10 V .

The number of electrons transferred is 2 and so we have:

$$\begin{aligned}
 E &= E^\ominus - \left(\frac{RT}{nF} \right) \ln Q \\
 1.06 &= 1.10 - \left(\frac{8.31 \times 298}{2 \times 96\,500} \right) \ln Q
 \end{aligned}$$

Rearranging this gives:

$$\ln Q = (1.10 - 1.06) \times \frac{2 \times 96\,500}{8.31 \times 298} = 3.12$$

To extract Q from this we need to raise both sides as powers of e :

$$Q = e^{3.12} = 22.6$$

Concentrations of solids are taken as 1. So:

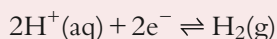
$$Q = \frac{[\text{Zn}^{2+}(\text{aq})]}{[\text{Cu}^{2+}(\text{aq})]}$$

$$[\text{Cu}^{2+}(\text{aq})] = \frac{[\text{Zn}^{2+}(\text{aq})]}{Q}$$

$$= \frac{1}{22.6} = 0.0443 \text{ mol dm}^{-3}$$

Equilibrium constants expressed in terms of partial pressures have not been covered anywhere else in the course, so it is extremely unlikely that you will get any questions involving this.

In a mixture of gases, the partial pressure of one of the component gases is the pressure that gas would exert if it were in the container by itself. This is its contribution to the total pressure of the mixture. Partial pressures are calculated relative to a standard state of 1 bar (100 kPa) and are inserted into the expression for the reaction quotient (Topic 7, page 287 in the Coursebook). So, if the partial pressure of hydrogen gas in a mixture is 80 kPa, this is equivalent to 0.8 bar. Consider the half-equation:



The standard electrode potential is 0.00 V but if the partial pressure of hydrogen is 0.800 bar, the temperature is 298 K and the concentration of H^+ ions is $0.100 \text{ mol dm}^{-3}$ we get:

$$\begin{aligned} Q &= \frac{P_{\text{H}_2}}{[\text{H}^+(\text{aq})]^2} \\ &= \frac{0.800}{0.100^2} \\ &= 80.0 \end{aligned}$$

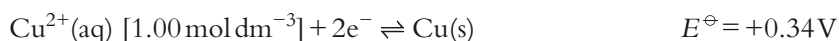
The number of electrons transferred is 2 and so we have:

$$\begin{aligned} E &= E^\ominus - \left(\frac{RT}{nF} \right) \ln Q \\ &= 0.00 - \left(\frac{8.31 \times 298}{2 \times 96\,500} \right) \ln 80 \\ &= -0.0562 \text{ V} \end{aligned}$$

Concentration cells

From the Nernst equation, we can see that an electrode potential depends on the concentration of the electrolyte and therefore it should be possible to set up a cell with two half-cells using the same chemicals but with different concentrations of electrolyte (Figure C.39). This is called a **concentration cell**.

The standard electrode potential for $\text{Cu}^{2+}(\text{aq})|\text{Cu}(\text{s})$ is 0.34 V and we saw earlier (page 53) that the electrode potential is 0.31 V when the concentration of $\text{Cu}^{2+}(\text{aq})$ is $0.100 \text{ mol dm}^{-3}$ and so we have:



and

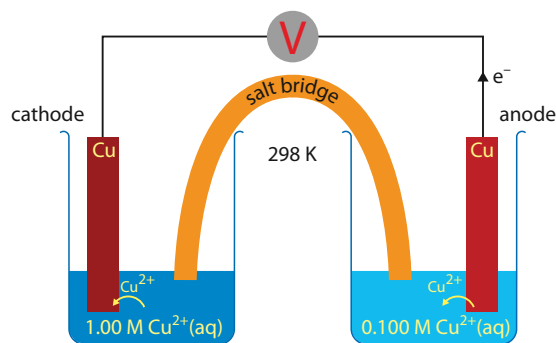
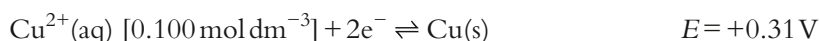
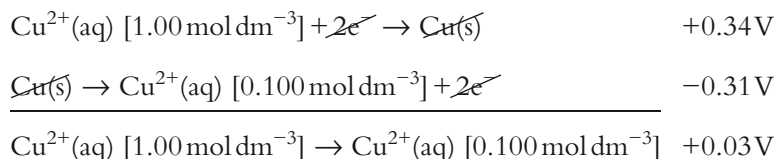


Figure C.39 A concentration cell.

A positive cell potential indicates a spontaneous change and so the half-equation involving the more negative (less positive) electrode potential is reversed to give the overall equation:



So, Cu^{2+} ions are reduced to copper in the left-hand half-cell and copper is oxidised to Cu^{2+} ions in the right-hand half-cell. Overall the reaction is equivalent to the concentration of Cu^{2+} ions in the left-hand half-cell decreasing and that in the right-hand half-cell increasing until they reach equilibrium – when the concentrations are equal and there is no overall cell potential.

The cell potential could also have been worked out directly from the Nernst equation. The standard electrode potential for the overall reaction is 0.00 V because there are the same species on both sides and, under standard conditions, the concentration of both would be 1.00 mol dm^{-3} so there would be no reaction.

The reaction quotient for the reaction is given by:

$$Q = \frac{0.100}{1.00} = 0.100$$

The number of electrons transferred is 2 and so we have:

$$\begin{aligned}
 E &= E^{\ominus} - \left(\frac{RT}{nF} \right) \ln Q \\
 &= 0.00 - \left(\frac{8.31 \times 298}{2 \times 96\,500} \right) \ln 0.100 \\
 &= 0.0300\text{V}
 \end{aligned}$$

Worked example

C.10 Calculate the cell potential for the concentration cell

$\text{Pt(s)} | \text{Co}^{2+}(\text{aq}) [0.500 \text{ mol dm}^{-3}], \text{Co}^{3+}(\text{aq}) [0.0500 \text{ mol dm}^{-3}] || \text{Co}^{3+}(\text{aq}) [0.600 \text{ mol dm}^{-3}],$
 $| \text{Co}^{2+}(\text{aq}) [0.0100 \text{ mol dm}^{-3}] | \text{Pt(s)}$. The standard electrode potential for this system is 1.82 V.

First of all write down the reduction half-equations:



$$Q = \frac{[\text{Co}^{2+}(\text{aq})]}{[\text{Co}^{3+}(\text{aq})]} \text{ for both reactions and } n = 1.$$

The electrode potential of each is then calculated using the Nernst equation.

$$\text{For reaction 1, } Q = \frac{0.500}{0.0500} = 10.0$$

$$\begin{aligned}
 E &= E^{\ominus} - \left(\frac{RT}{nF} \right) \ln Q \\
 &= 1.82 - \left(\frac{8.31 \times 298}{1 \times 96\,500} \right) \ln 10.0 \\
 &= 1.76 \text{ V}
 \end{aligned}$$

For reaction 2, $Q = \frac{0.0100}{0.600} = 0.0167$

$$\begin{aligned}
 E &= E^{\ominus} - \left(\frac{RT}{nF} \right) \ln Q \\
 &= 1.82 - \left(\frac{8.31 \times 298}{1 \times 96\,500} \right) \ln 0.0167 \\
 &= 1.93 \text{ V}
 \end{aligned}$$

For $\text{Co}^{3+}(\text{aq}) [0.0500 \text{ mol dm}^{-3}] + \text{e}^{-} \rightleftharpoons \text{Co}^{2+}(\text{aq}) [0.500 \text{ mol dm}^{-3}]$ $E = 1.76 \text{ V}$

For $\text{Co}^{3+}(\text{aq}) [0.600 \text{ mol dm}^{-3}] + \text{e}^{-} \rightleftharpoons \text{Co}^{2+}(\text{aq}) [0.0100 \text{ mol dm}^{-3}]$ $E = 1.93 \text{ V}$

The half-equation with the more negative (less positive) electrode potential is reversed so the cell potential is given by:

$$E_{\text{cell}} = -1.76 + 1.93 = 0.17 \text{ V}$$

Nature of science

Research into batteries is one of those fields where the distinction between science and technology is blurred. Developments in science and technology often go hand-in-hand and the development of better batteries has allowed the design and use of more and more sophisticated electronic devices and changed society.

Advances in science and technology have often been associated with environmental problems. Batteries can contain heavy metals and should be recycled rather than dumped in landfill sites, which can lead to pollution of soil and water with heavy metals. Many countries have laws governing the recycling of batteries.

Test yourself

- 21 Calculate the electrode potential for each of the following half-cells at 298 K. Standard electrode potentials are given in the table.

Half-cell	E^\ominus / V
$\text{Pb}^{2+}(\text{aq}) + 2\text{e}^- \rightleftharpoons \text{Pb}(\text{s})$	-0.13
$\text{Ag}^+(\text{aq}) + \text{e}^- \rightleftharpoons \text{Ag}(\text{s})$	+0.80
$\text{Fe}^{3+}(\text{aq}) + \text{e}^- \rightleftharpoons \text{Fe}^{2+}(\text{aq})$	+0.77
$\text{Co}^{3+}(\text{aq}) + \text{e}^- \rightleftharpoons \text{Co}^{2+}(\text{aq})$	+1.82
$\text{Cr}^{3+}(\text{aq}) + 3\text{e}^- \rightleftharpoons \text{Cr}(\text{s})$	-0.74
$\text{Cu}^{2+}(\text{aq}) + 2\text{e}^- \rightleftharpoons \text{Cu}(\text{s})$	+0.34

- a $\text{Pb}^{2+}(\text{aq}) + 2\text{e}^- \rightleftharpoons \text{Pb}(\text{s})$ when the concentration of $\text{Pb}^{2+}(\text{aq})$ is $0.0500 \text{ mol dm}^{-3}$
- b $\text{Ag}^+(\text{aq})|\text{Ag}(\text{s})$ when the concentration of $\text{Ag}^+(\text{aq})$ is $0.0100 \text{ mol dm}^{-3}$
- c $\text{Fe}^{3+}(\text{aq})|\text{Fe}^{2+}(\text{aq})$ when the concentration of $\text{Fe}^{3+}(\text{aq})$ is $0.0800 \text{ mol dm}^{-3}$ and that of $\text{Fe}^{2+}(\text{aq})$ is $0.0200 \text{ mol dm}^{-3}$
- d $\text{Co}^{3+}(\text{aq})|\text{Co}^{2+}(\text{aq})$ when the concentration of $\text{Co}^{3+}(\text{aq})$ is $0.600 \text{ mol dm}^{-3}$ and that of $\text{Co}^{2+}(\text{aq})$ is $0.600 \text{ mol dm}^{-3}$
- e $\text{Cr}^{3+}(\text{aq}) + 3\text{e}^- \rightleftharpoons \text{Cr}(\text{s})$ when the concentration of $\text{Cr}^{3+}(\text{aq})$ is $0.0200 \text{ mol dm}^{-3}$
- 22 Calculate the electrode potential for:
 $\text{MnO}_4^-(\text{aq}) + 8\text{H}^+(\text{aq}) + 5\text{e}^- \rightleftharpoons \text{Mn}^{2+}(\text{aq}) + 4\text{H}_2\text{O}(\text{l}) \quad E^\ominus = 1.51 \text{ V}$
 under the following conditions – the temperature is 298 K in each case.
- a $[\text{MnO}_4^-(\text{aq})] = 0.0200 \text{ mol dm}^{-3}$, $[\text{Mn}^{2+}(\text{aq})] = 0.0500 \text{ mol dm}^{-3}$, $\text{pH} = 3.50$
- b $[\text{MnO}_4^-(\text{aq})] = 0.00100 \text{ mol dm}^{-3}$, $[\text{Mn}^{2+}(\text{aq})] = 0.00200 \text{ mol dm}^{-3}$, $\text{pH} = 2.00$
- c $[\text{MnO}_4^-(\text{aq})] = 0.100 \text{ mol dm}^{-3}$, $[\text{Mn}^{2+}(\text{aq})] = 0.0800 \text{ mol dm}^{-3}$, $\text{pH} = 1.70$
- 23 Calculate the cell potentials for each of the following voltaic cells under the stated conditions and 298 K. You will also need to use the values given in the table in question 21.
- a $\text{Cu}(\text{s}) + 2\text{Ag}^+(\text{aq}) \rightarrow \text{Cu}^{2+}(\text{aq}) + 2\text{Ag}(\text{s})$, where $[\text{Ag}^+(\text{aq})] = 0.0100 \text{ mol dm}^{-3}$ and $[\text{Cu}^{2+}(\text{aq})] = 0.0200 \text{ mol dm}^{-3}$
- b $\text{Co}^{3+}(\text{aq}) + \text{Ag}(\text{s}) \rightarrow \text{Co}^{2+}(\text{aq}) + \text{Ag}^+(\text{aq})$, where $[\text{Co}^{3+}(\text{aq})] = 0.300 \text{ mol dm}^{-3}$, $[\text{Co}^{2+}(\text{aq})] = 0.500 \text{ mol dm}^{-3}$ and $[\text{Ag}^+(\text{aq})] = 0.100 \text{ mol dm}^{-3}$
- c $2\text{Cr}(\text{s}) + 3\text{Pb}^{2+}(\text{aq}) \rightarrow 2\text{Cr}^{3+}(\text{aq}) + 3\text{Pb}(\text{s})$, where $[\text{Pb}^{2+}(\text{aq})] = 0.0100 \text{ mol dm}^{-3}$ and $[\text{Cr}^{3+}(\text{aq})] = 0.0500 \text{ mol dm}^{-3}$
- 24 Calculate the cell potential for each of the following concentration cells at 298 K. Use the data in the table in question 21.
- a $\text{Ag}(\text{s})|\text{Ag}^+(\text{aq})[0.100 \text{ mol dm}^{-3}]||\text{Ag}^+(\text{aq})[1.00 \text{ mol dm}^{-3}]\text{Ag}(\text{s})$
- b $\text{Cr}(\text{s})|\text{Cr}^{3+}(\text{aq})[0.0200 \text{ mol dm}^{-3}]||\text{Cr}^{3+}(\text{aq})[1.00 \text{ mol dm}^{-3}]\text{Cr}(\text{s})$
- c $\text{Cu}(\text{s})|\text{Cu}^{2+}(\text{aq})[0.0100 \text{ mol dm}^{-3}]||\text{Cu}^{2+}(\text{aq})[0.300 \text{ mol dm}^{-3}]\text{Cu}(\text{s})$
- d $\text{Pt}(\text{s})|\text{Fe}^{2+}(\text{aq})[0.800 \text{ mol dm}^{-3}], \text{Fe}^{3+}(\text{aq})[0.200 \text{ mol dm}^{-3}]||\text{Fe}^{3+}(\text{aq})[0.180 \text{ mol dm}^{-3}], \text{Fe}^{2+}(\text{aq})[0.0200 \text{ mol dm}^{-3}]\text{Pt}(\text{s})$

C7 Nuclear fusion and nuclear fission (HL)

C7.1 Nuclear binding energy

When protons and neutrons come together to form a nucleus, the total mass of the nucleus is less than the mass of the protons and neutrons from which it was made. This difference in mass is called the mass defect (Δm).

Mass defect (Δm) is the difference between the mass of a nucleus and the sum of the masses of the individual nucleons.

This mass is converted to energy when the nucleus is formed according to Einstein's equation:

$$E = mc^2$$

where E is energy (in J), m is the mass (in kg) and c is the speed of light ($3.00 \times 10^8 \text{ ms}^{-1}$).

The nuclear binding energy (ΔE) is the energy required to break apart a nucleus into individual protons and neutrons.

The binding energy is not something that the nucleus 'has' – it is the energy released when the nucleus is formed, or that is required to break it apart again. If the binding energy for a nucleus is divided by the total number of nucleons we get the average binding energy per nucleon, which is a measure of the stability of the nucleus.

The mass of subatomic particles is often expressed in terms of the **unified atomic mass unit** (u), which is $\frac{1}{12}$ of the mass of a carbon-12 atom. The mass of one mole of carbon-12 atoms is 0.01200 kg and, therefore, the mass of a carbon atom can be worked out by dividing this by Avogadro's constant:

$$\frac{0.01200}{6.02 \times 10^{23}} = 1.99 \times 10^{-26} \text{ kg}$$

The mass equivalent to 1 u is $\frac{1}{12}$ of this:

$$\frac{1}{12} \times 1.99 \times 10^{-26} = 1.66 \times 10^{-27} \text{ kg}$$

Therefore 1 u is equivalent to $1.66 \times 10^{-27} \text{ kg}$. We will use this value in all later work.

The masses of protons, neutrons and electrons are given in Table C.12.

Particle	Mass	Mass
proton	$1.672622 \times 10^{-27} \text{ kg}$	1.007276 u
neutron	$1.674927 \times 10^{-27} \text{ kg}$	1.008665 u
electron	$9.109383 \times 10^{-31} \text{ kg}$	0.0005486 u

Table C.12 Masses of subatomic particles.

Learning objectives

- Understand what is meant by the terms mass defect and nuclear binding energy
- Calculate the mass defect and nuclear binding energy of a nucleus
- Calculate the energy released in fission and fusion reactions

The binding energy could also be defined in terms of the energy released when the nucleons come together to form a particular nucleus.

IUPAC gives the value of 1 u as $1.6605402 \times 10^{-27} \text{ kg}$.

A mass defect can be worked out as follows, using a helium isotope as an example. The mass of a ${}^4_2\text{He}$ atom is 4.002 60 u. If we subtract the mass of the two electrons we get the mass of the nucleus:

$$\text{mass of the nucleus} = 4.002\,60 - (2 \times 0.000\,548\,6) = 4.001\,502\,8\,\text{u}$$

This nucleus is made up of two protons and two neutrons. The mass defect is the difference between the mass of the nucleus and the masses of the individual nucleons:

$$\begin{aligned}\Delta m &= (2 \times 1.007\,276) + (2 \times 1.008\,665) - 4.001\,508 \\ &= 0.030\,374\,\text{u}\end{aligned}$$

This can be converted to kilograms by multiplying by 1.66×10^{-27} :

$$\begin{aligned}\text{mass defect} &= 0.030\,374 \times (1.66 \times 10^{-27}) \\ &= 5.04 \times 10^{-29}\,\text{kg}\end{aligned}$$

This mass is converted into energy when the nucleus is formed and so the binding energy can be worked out using the equation $E=mc^2$:

$$\Delta E = (5.04 \times 10^{-29}) \times (3.00 \times 10^8)^2 = 4.54 \times 10^{-12}\,\text{J}$$

This represents the energy required to break apart an ${}^4_2\text{He}$ nucleus into its individual nucleons, or the energy released when the nucleus is formed from the individual nucleons. This may not seem very much but when we work it out for one mole of helium, we get:

$$(4.54 \times 10^{-12}) \times (6.02 \times 10^{23}) = 2.73 \times 10^{12}\,\text{J mol}^{-1}, \text{ or } 2.73 \times 10^9\,\text{kJ mol}^{-1}$$

This can be compared to the amount of energy released when one mole of carbon is burned, which is $394\,\text{kJ mol}^{-1}$.

The binding energy per nucleon can be worked out by dividing the binding energy by the number of nucleons, four in this case:

$$\text{binding energy per nucleon} = \frac{4.54 \times 10^{-12}}{4} = 1.14 \times 10^{-12}\,\text{J}$$

These energies are often expressed in more convenient units such as MeV (megaelectronvolts), where $1\,\text{MeV}$ is $1.602 \times 10^{-13}\,\text{J}$.

$$\text{binding energy per nucleon for } {}^4_2\text{He} = \frac{1.14 \times 10^{-12}}{1.602 \times 10^{-13}} = 7.12\,\text{MeV}$$

Exam tip

The mass must be in kilograms in this equation.

If more figures are carried through on the calculator, the values for binding energy per nucleon come out as $1.13 \times 10^{-12}\,\text{J}$ and $7.08\,\text{MeV}$.



Test yourself

- 25 Calculate the mass defect (in u), the binding energy (in J) and the binding energy per nucleon (in MeV) for each of the following nuclei:
- a ${}^{10}_5\text{B}$ (mass 10.012 937 u)
 - b ${}^{23}_{11}\text{Na}$ (mass 22.989 770 u)
 - c ${}^{56}_{26}\text{Fe}$ (mass 55.934 942 u)
 - d ${}^{239}_{94}\text{Pu}$ (mass 239.052 157 u)

Calculating the energy released in fission and fusion reactions

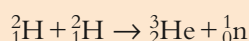
Use the values given in Table C.13 when working through the worked examples that follow.

Atom	Mass / u
^1_1H	1.007 825
^2_1H	2.014 102
^3_1H	3.016 049
^3_2He	3.016 029
^4_2He	4.002 60
$^{141}_{56}\text{Ba}$	140.914 411
$^{92}_{36}\text{Kr}$	91.926 156
$^{235}_{92}\text{U}$	235.043 923

Table C.13 The masses of some atoms in unified atomic mass units.

Worked examples

C.11 Given the atomic masses in Table C.13 and the mass of a neutron given in Table C.12, calculate the energy released in kJ mol^{-1} in the fusion reaction:



The change in mass (Δm) is calculated by finding the difference between the mass of reactants and the mass of the products:

$$\text{total mass of reactants} = 2 \times 2.014\,102 = 4.028\,204\,\text{u}$$

$$\text{total mass of products} = 3.016\,029 + 1.008\,665 = 4.024\,694\,\text{u}$$

It can be seen that the total mass of the products is less than the total mass of the reactants:

$$\Delta m = 4.024\,694 - 4.028\,204 = -0.003\,51\,\text{u}$$

The mass change can be converted to kg by multiplying by 1.66×10^{-27} :

$$\Delta m = 0.003\,51 \times (1.66 \times 10^{-27}) = 5.83 \times 10^{-30}\,\text{kg}$$

This mass is converted to energy when the fusion reaction occurs.

The energy released is worked out by using the Einstein equation:

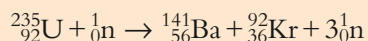
$$\begin{aligned} E &= mc^2 \\ &= (5.83 \times 10^{-30}) \times (3.00 \times 10^8)^2 \\ &= 5.25 \times 10^{-13}\,\text{J} \end{aligned}$$

The energy released per mole is:

$$\begin{aligned} E &= (5.25 \times 10^{-13}) \times (6.02 \times 10^{23}) \\ &= 3.16 \times 10^{11}\,\text{J mol}^{-1} \text{ or } 3.16 \times 10^8\,\text{kJ mol}^{-1} \end{aligned}$$

This is the energy released when two moles of deuterium atoms undergo fusion to form ^3He and it is more than a million times the amount of energy released when one mole of H_2 , which contains the same number of atoms, is burned.

C.12 Given the atomic masses in Table C.13 and the mass of a neutron in Table C.12, calculate the energy released in kJ mol^{-1} in the fission reaction:



This calculation is very similar to the previous one. The change in mass is calculated by working out the difference between the mass of reactants and the mass of the products:

$$\text{total mass of reactants} = 235.043\,923 + 1.008\,665 = 236.052\,588 \text{ u}$$

$$\text{total mass of products} = 140.914\,411 + 91.926\,156 + (3 \times 1.008\,665) = 235.866\,562 \text{ u}$$

$$\Delta m = 235.866\,562 - 236.052\,588$$

$$= -0.186\,026 \text{ u}$$

The mass change can be converted to kg by multiplying by 1.66×10^{-27} :

$$\text{mass change} = 0.186\,026 \times 1.66 \times 10^{-27}$$

$$= 3.09 \times 10^{-28} \text{ kg}$$

This mass is converted into energy in the fission reaction and the energy released is worked out using the Einstein equation:

$$E = mc^2$$

$$= (3.09 \times 10^{-28}) \times (3.00 \times 10^8)^2$$

$$= 2.78 \times 10^{-11} \text{ J}$$

The energy released per mole is:

$$E = (2.78 \times 10^{-11}) \times (6.02 \times 10^{23})$$

$$= 1.67 \times 10^{13} \text{ J mol}^{-1}, \text{ or } 1.67 \times 10^{10} \text{ kJ mol}^{-1}$$

From this we could work out that the energy released when 1 g of uranium-235 undergoes fission is

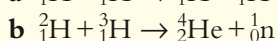
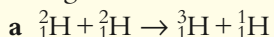
$$\frac{1.67}{235} \times 10^{10} = 7.11 \times 10^7 \text{ kJ g}^{-1}. \text{ This can be compared}$$

with the specific energy of octane, which is only 47.9 kJ g^{-1} .

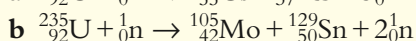
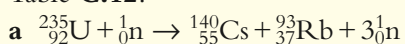
The electrons have been included in all of the calculations above but the calculations are sometimes carried out just using the masses of nuclei (obtained by subtracting the total mass of the electrons from the mass of the atom). However, since there is the same number of electrons on both sides of the equation, their masses will cancel out and make no difference to the overall calculation.

? Test yourself

26 Using the data in Table C.13 calculate the energy released (in kJ mol^{-1}) in each of these fusion reactions:



27 Calculate the energy released (in kJ mol^{-1}) in each of the following fission reactions using the data in the table below and the mass of a neutron given in Table C.12.



Atom	Mass / u
${}^{140}\text{Cs}$	139.91728
${}^{93}\text{Rb}$	92.92204
${}^{105}\text{Mo}$	104.9170
${}^{129}\text{Sn}$	128.9135
${}^{235}\text{U}$	235.043923

C7.2 Half-life

Radioactive decay is a first-order process, so the rate of decay is proportional to the number of undecayed nuclei remaining. We normally discuss the rate of decay in terms of the activity (A), which is the number of nuclei which decay per second, and so we can write a rate equation for radioactive decay as:

$$A = \lambda N$$

where λ is the decay constant and N is the number of undecayed nuclei present. The unit of activity is the becquerel (Bq), which is equivalent to 1 disintegration (decay) per second.

This can be compared to a first-order rate equation for a chemical reaction:

$$\text{rate} = k[X]$$

as described in Topic 6. It can be seen that the decay constant is equivalent to a rate constant.

Integration of the rate equation for radioactive decay in calculus form produces the equation:

$$N = N_0 e^{-\lambda t}$$

where N_0 is the initial number of undecayed nuclei present and N is the number of undecayed nuclei present at time t .

Further mathematical manipulation produces an equation that relates the decay constant to the half-life:

$$\lambda = \frac{\ln 2}{t_{\frac{1}{2}}}$$

It can be seen from this equation that, like the first-order rate constant, the units of the decay constant are time^{-1} . If the activity is measured in becquerels then the decay constant should have units of s^{-1} .

Learning objectives

- Understand that radioactive decay is a first order process
- Use the relationship between the decay constant and half-life
- Solve problems involving non-integral numbers of half-lives

The rate equation for radioactive decay using calculus notation is:

$$\frac{dN}{dt} = -\lambda N$$

This equation could also be written in the form

$$\ln\left(\frac{N}{N_0}\right) = -\lambda t$$

Worked examples

C.13 The half-life of rutherfordium-104 is 65 s. Calculate the decay constant and the percentage of a $1.00 \mu\text{g}$ sample remaining after 3.00 minutes.

$$\begin{aligned}\lambda &= \frac{\ln 2}{t_{\frac{1}{2}}} \\ &= \frac{0.693}{65} = 0.011 \text{ s}^{-1}\end{aligned}$$

This answer has been rounded to two significant figures but more figures have been carried through for further calculations.

Next we need to use $N = N_0 e^{-\lambda t}$. Both λ and t must be in the same units of time – so if λ is in s^{-1} , t must be in seconds. 3.00 minutes is 180 s.

We are trying to find the ratio $\frac{N}{N_0}$, and we have:

$$\frac{N}{N_0} = e^{-\lambda t} = e^{-0.011 \times 180} = 0.147$$

This is multiplied by 100 to get a percentage and so about 15% remains undecayed after 3.00 minutes.

C.14 Given that the activity of a $1.00\text{ }\mu\text{g}$ sample of nitrogen-13 is $5.35 \times 10^{13}\text{ Bq}$, calculate the half-life and the mass left after 45 minutes. The mass of a nitrogen-13 atom is $2.16 \times 10^{-23}\text{ g}$.

The number of nitrogen-13 atoms in $1.00\text{ }\mu\text{g}$ is:

$$\frac{1.00 \times 10^{-6}}{2.16 \times 10^{-23}} \text{ i.e. } 4.63 \times 10^{16} \text{ atoms}$$

$A = \lambda N$ so:

$$\begin{aligned} \lambda &= \frac{A}{N} \\ &= \frac{5.35 \times 10^{13}}{4.63 \times 10^{16}} = 1.16 \times 10^{-3}\text{ s}^{-1} \end{aligned}$$

$$\begin{aligned} t_{\frac{1}{2}} &= \frac{\ln 2}{\lambda} \\ &= \frac{0.693}{1.16 \times 10^{-3}} = 600\text{ s} \end{aligned}$$

To calculate the mass left after 45 minutes we use:

$$N = N_0 e^{-\lambda t} \text{ in the form } \frac{N}{N_0} = e^{-\lambda t}$$

The decay constant is in units of s^{-1} so we must convert the time to seconds:

$$t = 45 \times 60 = 2700\text{ s}$$

$$\begin{aligned} \frac{N}{N_0} &= e^{-\lambda t} \\ &= e^{-1.16 \times 10^{-3} \times 2700} \\ &= 0.0442 \end{aligned}$$

This is the proportion of the sample *undecayed* after 45 minutes, so the mass of nitrogen-13 left at the end is 0.0442 times the original mass:

$$\begin{aligned} \text{mass} &= 0.0442 \times 1.00 \\ &= 0.0442\text{ }\mu\text{g, or } 4.42 \times 10^{-8}\text{ g} \end{aligned}$$

It is important to be consistent with units in these questions. The decay constant in the above example could also have been written as 0.0693 min^{-1} . This is just 60 times the decay constant given.

Using the decay constant in this form means that time can be used in minutes. Substituting in $\frac{N}{N_0} = e^{-\lambda t}$ we get:

$$\frac{N}{N_0} = e^{-0.0693 \times 45}, \text{ which gives the same answer as above.}$$

C.15 1.00 kg of a particular rock was believed to contain 0.120 g of potassium-40 when it was originally formed. Analysis of the rock has determined that it now contains 0.0500 g of potassium-40. Calculate the age of the rock given that the half-life of potassium-40 is 1.25×10^9 years.

We will need to use $\frac{N}{N_0} = e^{-\lambda t}$ and so must first work out the decay constant:

$$\begin{aligned}\lambda &= \frac{\ln 2}{t_{\frac{1}{2}}} \\ &= \frac{0.693}{1.25 \times 10^9} \\ &= 5.54 \times 10^{-10} \text{ y}^{-1}\end{aligned}$$

We can use mass instead of number of atoms in the following equation because any conversion factor will simply cancel out in the ratio.

$$\begin{aligned}\frac{N}{N_0} &= e^{-\lambda t} \\ \frac{0.0500}{0.120} &= e^{-5.54 \times 10^{-10} t}\end{aligned}$$

$$\text{So } e^{-5.54 \times 10^{-10} t} = 0.4167$$

Taking the natural log of both sides:

$$-5.54 \times 10^{-10} t = \ln 0.4167$$

$$-5.54 \times 10^{-10} t = -0.875$$

$$\text{So } t = \frac{0.875}{5.54 \times 10^{-10}}$$

$$= 1.58 \times 10^9 \text{ y}$$

The decay constant was expressed in y^{-1} so the time will come out in years.

Exam tip

$\frac{N}{N_0} = e^{-\lambda t}$ can be used more conveniently in the form $\ln\left(\frac{N}{N_0}\right) = -\lambda t$, which makes the mathematical manipulation simpler in this type of problem.

Exam tip

As a final check we can consider if this answer seems reasonable. The half-life of potassium-40 is 1.25×10^9 years, therefore 0.120 g should decay to 0.0600 g in this time. We are trying to find the time that it takes to decay to 0.0500 g and therefore would expect it to be slightly longer than one half-life, which our answer is.

? Test yourself

- 28** Calculate the decay constant in each of the following cases:
 - a** the half-life of nobelium-259 is 58 minutes
 - b** the half-life of rubidium-83 is 86.2 days
 - c** the half-life of neodymium-144 is 2.1×10^{15} years
- 29** Assuming that you start in each case with $5.00 \mu\text{g}$ of the isotopes in question **28**, calculate how long it will take for the amount of the isotope to fall to $1.00 \mu\text{g}$.
- 30** Calculate what percentage of the stated isotope is left after the stated time.
 - a** carbon-14 (half-life 5730 years) after 10 000 years
 - b** iodine-131 (half-life 8.04 days) after 3 weeks
- 31** Given that activity of a $2.00 \mu\text{g}$ sample of osmium-191 is $3.27 \times 10^9 \text{ Bq}$, calculate the half-life and the mass left after 6 weeks. The mass of an osmium-191 atom is $3.19 \times 10^{-22} \text{ g}$.

Learning objectives

- Understand that the rate of effusion of a gas is inversely proportional to the square root of its molar mass
- Understand Graham's law of effusion in terms of the kinetic theory
- Solve problems involving Graham's law of effusion
- Understand the different methods of uranium enrichment
- Describe the different physical properties of UO_2 and UF_6 in terms of structure and bonding

Graham's law is often stated in terms of the density of the gas – the rate of effusion is inversely proportional to the density. The density of a gas is proportional to its relative molecular mass at a particular temperature and pressure.

Basically what we are saying here is that, because average kinetic energy is the same for all gases at the same temperature, heavier molecules move more slowly and therefore will leave the container more slowly.

C7.3 Enrichment of uranium

Graham's law of effusion

Effusion is the process by which a gas escapes through a very small hole in a container (Figure C.40).

Graham's law: the rate of effusion of a gas is inversely proportional to the relative molecular mass of the gas.

The origin of Graham's law can be explained in terms of the kinetic theory. The average kinetic energy of the particles in a gas is proportional to the absolute temperature:

$$KE \propto T$$

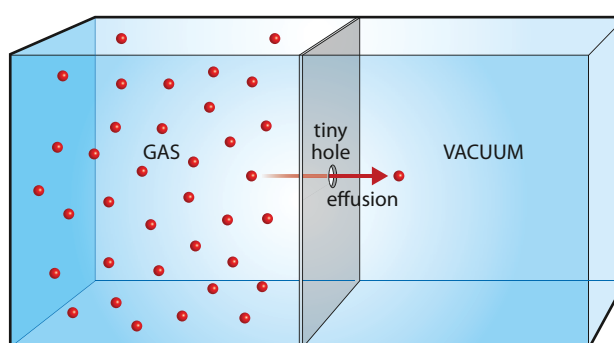


Figure C.40 Gas molecules escape through a very small hole in the container.

The average kinetic energy is given by $\frac{1}{2}m\bar{v}^2$, where m is the mass of a molecule and \bar{v}^2 represents the mean square speed – the average of the squared speeds.

$$\text{So } \frac{1}{2}m\bar{v}^2 \propto T$$

$$\text{and } \bar{v}^2 \propto \frac{T}{m}$$

So at a constant temperature $\bar{v} \propto 1/\sqrt{m}$, where \bar{v} is the root mean square speed which is very similar to the average speed of the molecules in the sample of gas.

How quickly a gas effuses through a small hole will depend on how often gas molecules 'hit' the hole, and this depends on the speed of the gas molecules. From this we can deduce that the rate of effusion is inversely proportional to the square root of the mass of a molecule – in other words, inversely proportional to the square root of the molar mass of the gas.

For two different gases in a container, Graham's law of effusion can be stated in the form:

$$\frac{\text{rate of effusion of gas 1}}{\text{rate of effusion of gas 2}} = \frac{\sqrt{\text{molar mass of gas 2}}}{\sqrt{\text{molar mass of gas 1}}}$$

$$\frac{\text{rate}_1}{\text{rate}_2} = \sqrt{\frac{M_2}{M_1}}$$

This assumes that if the gases are in the same container they will be at the same temperature and pressure.

Worked examples

C.16 Equal numbers of moles of carbon dioxide and oxygen are put in a container containing a small hole in its wall. What is the ratio of the rate of effusion of carbon dioxide to that of oxygen? After one hour will there be more moles of carbon dioxide or of oxygen in the container?

Molar mass of $\text{CO}_2 = 44.01 \text{ g mol}^{-1}$; molar mass of $\text{O}_2 = 32.00 \text{ g mol}^{-1}$

$$\begin{aligned}\frac{\text{rate of effusion of CO}_2}{\text{rate of effusion of O}_2} &= \frac{\sqrt{(\text{molar mass of O}_2)}}{\sqrt{(\text{molar mass of CO}_2)}} \\ &= \frac{\sqrt{32.00}}{\sqrt{44.01}} \\ &= 0.853\end{aligned}$$

This means that oxygen effuses 1.173 times ($\frac{1}{0.853}$) more quickly than carbon dioxide and, because there were equal numbers of molecules present originally, after one hour more moles of carbon dioxide will be in the container.

C.17 A gas, X, effuses through a small hole at the rate of $1.20 \text{ cm}^3 \text{ h}^{-1}$. Under the same conditions of temperature and pressure, nitrogen gas effuses through the same hole at the rate of $1.54 \text{ cm}^3 \text{ h}^{-1}$. Calculate the molar mass of X.

Molar mass of $\text{N}_2 = 28.02 \text{ g mol}^{-1}$

$$\frac{\text{rate of effusion of N}_2}{\text{rate of effusion of X}} = \frac{\sqrt{(\text{molar mass of X})}}{\sqrt{(\text{molar mass of N}_2)}}$$

$$\frac{1.54}{1.20} = \frac{\sqrt{M_x}}{\sqrt{28.02}}$$

$$\text{So } \sqrt{M_x} = \frac{\sqrt{28.02} \times 1.54}{1.20}$$

$$\text{and } M_x = 46.1 \text{ g mol}^{-1}$$

C.18 A container of volume 1.00 dm^3 contains 0.0200 mol of oxygen and 0.0200 mol of helium.

The rate of effusion of oxygen through a very small hole in the container wall is $5.60 \times 10^{-5} \text{ mol h}^{-1}$. Calculate the rate of effusion of helium and the number of moles of each gas present in the container after 48 hours.

Molar mass of $\text{He} = 4.00 \text{ g mol}^{-1}$; molar mass of $\text{O}_2 = 32.00 \text{ g mol}^{-1}$

$$\begin{aligned}\frac{\text{rate of effusion of He}}{\text{rate of effusion of O}_2} &= \frac{\sqrt{(\text{molar mass of O}_2)}}{\sqrt{(\text{molar mass of He})}} \\ &= \frac{\sqrt{32.00}}{\sqrt{4.00}} \\ &= 2.83\end{aligned}$$

So helium effuses 2.83 times as quickly as oxygen. The rate of effusion of helium is $2.83 \times 5.60 \times 10^{-5} \text{ mol h}^{-1}$, or $1.58 \times 10^{-4} \text{ mol h}^{-1}$.

The amount of oxygen that will effuse in 48 hours is $48 \times 5.60 \times 10^{-5} = 2.69 \times 10^{-3} \text{ mol}$

The amount of oxygen remaining in the container will be $0.0200 - 2.69 \times 10^{-3} = 0.0173 \text{ mol}$

The amount of helium that will effuse in 48 hours is $48 \times 1.58 \times 10^{-4} = 7.60 \times 10^{-3} \text{ mol}$

The amount of helium remaining in the container will be $0.0200 - 7.60 \times 10^{-3} = 0.0124 \text{ mol}$

? Test yourself

- 32** Calculate the relative rates of effusion of the gas pairs below:
- helium (He) and sulfur dioxide (SO₂)
 - ethene (C₂H₄) and propane (C₃H₈)
 - hydrogen (H₂) and methane (CH₄)
- 33** Calculate the molar mass of the unknown gas in each case:
- nitrogen effuses at 1.07 times the rate of gas **Y**
 - gas **Q** effuses at 1.30 times the rate of carbon dioxide
 - helium effuses at 2.60 times the rate of gas **Z**
- 34** Mixtures of two gases are put into three different containers so that each container contains 5.00×10^{-3} mol of the gases listed below. In each case, the rate of effusion of one of the gases is given. Calculate the number of moles of each gas present after 24 hours:
- methane (rate of effusion 4.80×10^{-5} mol h⁻¹) and helium
 - carbon monoxide (rate of effusion 3.60×10^{-5} mol h⁻¹) and carbon dioxide
 - fluorine (rate of effusion 6.00×10^{-5} mol h⁻¹) and chlorine (assume no reaction between the gases under these conditions)

Enrichment requires large amounts of electricity and contributes significantly to the cost of nuclear power. Depending on how the electricity for the enrichment plant is generated, it can also make a significant difference to the carbon footprint of nuclear power generation.

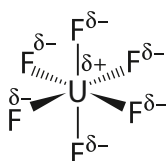


Figure C.41 The structure of UF₆.

The molecules are essentially spherical with a small area of contact between adjacent molecules and the fluorine atoms are not very polarisable – so the London forces between molecules are much weaker than in, for instance, a linear hydrocarbon of similar relative molecular mass.

Uranium enrichment

There are two main naturally occurring isotopes of uranium – uranium-235 and uranium-238. The isotope that undergoes fission in a nuclear reactor is uranium-235, but the problem is that the natural abundance of this isotope is only 0.72%. The proportion of uranium-235 present in a sample, therefore, generally has to be increased before it can be used in a nuclear reactor. The process by which this is done is called **enrichment**.

Different reactors require uranium that has been enriched to different extents but most require the fuel to be enriched to contain at least 3% uranium-235. Nuclear weapons require uranium that has been enriched to contain at least 90% uranium-235.

There are two main techniques for uranium enrichment: **gaseous diffusion** and **gas centrifugation**. Before enrichment can occur, the uranium must be converted into a volatile form – uranium hexafluoride, UF₆. This is made from uranium ore in a series of steps. UF₆ exists in the gaseous state during the enrichment process and both enrichment techniques rely on the difference in mass between gaseous ²³⁵UF₆ and ²³⁸UF₆ molecules.

Properties of UF₆ and UO₂

Uranium(VI) fluoride (uranium hexafluoride, UF₆) is a white crystalline solid that sublimates at 64 °C. It consists of UF₆ octahedral molecules in the solid and gaseous states (Figure C.41). There are covalent bonds between the uranium atom and the fluorine atoms, and London forces between the molecules in the solid state.

Because fluorine is much more electronegative than uranium, all the bonds are very polar. However, the symmetry of the molecule means that the dipoles cancel so the molecule is non-polar overall. There are therefore relatively weak forces (London forces) between the molecules and UF₆ sublimates at a relatively low temperature.

Uranium(IV) oxide (uranium dioxide, UO₂) is a dark-brown, crystalline ionic solid that melts at over 2800 °C. The solid consists of U⁴⁺ ions and O²⁻ ions with very strong electrostatic forces between the ions which require a lot of energy to break.

Gaseous diffusion

In gaseous diffusion, the UF_6 is forced at high pressure through a container with walls made of a semi-permeable membrane (Figure C.42). Because $^{235}\text{UF}_6$ has a slightly lower relative molecular mass, it effuses slightly faster through the membrane and the stream of gas becomes slightly enriched with uranium-235.

The rate of effusion of $^{235}\text{UF}_6$ is only 1.004 times that of $^{238}\text{UF}_6$ because the molar masses are so similar. This means that this process must be repeated several hundred times to achieve an enrichment of even a few percent.

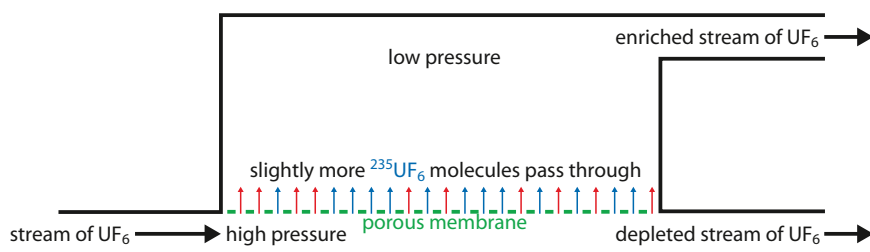


Figure C.42 Enrichment by gaseous diffusion.

Gas centrifugation

In gas centrifugation, the UF_6 molecules are fed into a cylinder (Figure C.43) which is rotated at high speed so that the slightly heavier $^{238}\text{UF}_6$ molecules move more towards the outside. The gas at the centre of the cylinder is slightly richer in the lighter $^{235}\text{UF}_6$ molecules. The slightly enriched stream then passes to the next centrifuge. As with diffusion, this must be repeated many times to achieve significant enrichment. Commercial plants operate large numbers of centrifuges in series and parallel.

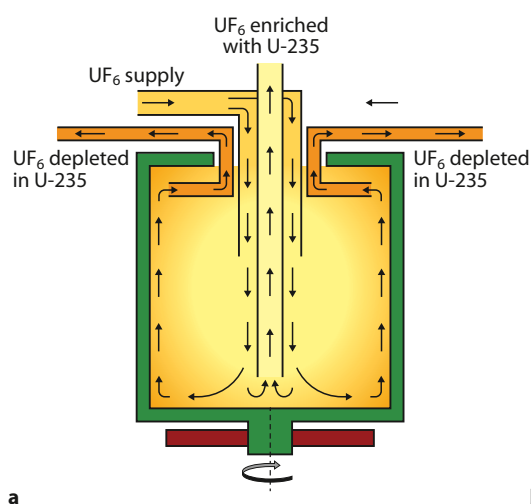


Figure C.43 a A gas centrifuge; b overhead view of rows of centrifuge units at the enrichment plant in Piketon, Ohio, USA.

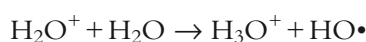
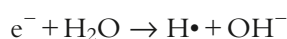
Learning objectives

- Understand that ionising radiation can cause damage to cells

C7.4 The dangers of radioactivity

The fuels used in nuclear reactors and the contents of nuclear waste are radioactive. They contain isotopes that undergo radioactive decay by the emission of alpha particles, beta particles and/or gamma rays (there are other forms of radioactive decay). These forms of radiation are called **ionising radiation** because they cause the formation of ions (by ejection of electrons) when they interact with matter.

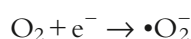
Ionising radiation can damage human cells and the main effect comes from damage caused to DNA. The ionising radiation can either interact directly with DNA, causing ionisation and a change of structure, or there can be indirect effects due to the formation of free radicals from other species, such as water. The most common substance in our body is water and when ionising radiation interacts with this the water molecules can become ionised. The ions and electrons generated can go on to react further to produce free radicals such as the hydroxyl radical ($\text{HO}\cdot$):



The last of these reactions could also be written $\text{H}_2\text{O}^+ \rightarrow \text{H}^+ + \text{HO}\cdot$.

Of the free radicals generated in these processes, the hydroxyl radical is probably the most dangerous. When it interacts with DNA it can set off a series of reactions that results in damage to the DNA. Free radicals can also cause damage to proteins (enzymes) and lipids in cells.

Other free radicals, such as the superoxide ion, can also be formed. This ion can be formed by the reaction of oxygen with an electron (liberated during an ionisation process):



The superoxide ion can be converted into hydroxyl radicals in cells – see Option A.

Nature of science

There have been many large-scale collaborative projects in the history of science and one of the most (in)famous is the Manhattan project, which resulted in the development of the first atomic bombs. At its peak it employed well over 100 000 workers. Two different bombs (one using uranium and the other plutonium) were developed and dropped on Hiroshima and Nagasaki in August 1945. There are many estimates as to how many people died as a result of this, and most are well over 100 000.

C8 Photovoltaic and dye-sensitised solar cells (HL)

C8.1 The effect of conjugation on the wavelength of light absorbed by molecules

Electromagnetic radiation in the visible–ultraviolet region of the spectrum is absorbed to promote an electron from a low energy level (molecular orbital) in a molecule to a higher energy level (molecular orbital). The wavelength at which the maximum amount of radiation is absorbed is called λ_{max} . We have already met (page 27) the idea that double bonds (chromophores) are needed if electromagnetic radiation is to be absorbed by a molecule and that a conjugated system is a system of alternating single and double bonds. We can see from Figure C.44 that electrons are delocalised in a conjugated system because p orbitals can overlap along the whole chain.

It can be seen from Table C.14 that:

the longer the conjugated chain (delocalised system), the longer the wavelength of radiation absorbed by a molecule.

As the conjugated system gets longer, the energy gap between the lower molecular orbital that the electron is promoted from and the higher molecular orbital that it is promoted to gets smaller (Figure C.45).

If the conjugated system is long enough, light in the visible region of the spectrum will be absorbed and the compound will be coloured.

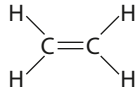
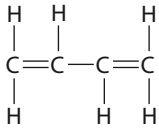
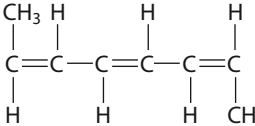
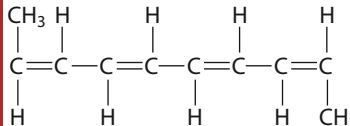
Molecule	No. of conjugated double bonds	$\lambda_{\text{max}} / \text{nm}$
	1	162
	2	217
	3	275
	4	310

Table C.14 The effect of conjugation on the wavelength of electromagnetic radiation absorbed.

Learning objectives

- Understand that molecules with longer conjugated systems absorb longer wavelength electromagnetic radiation

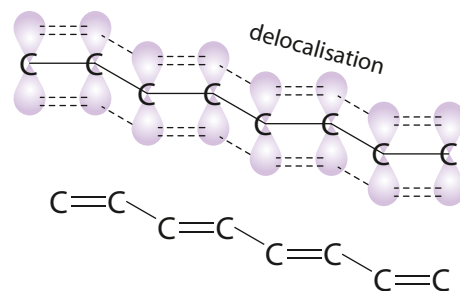


Figure C.44 Delocalised electrons and overlapping p orbitals in a conjugated system.

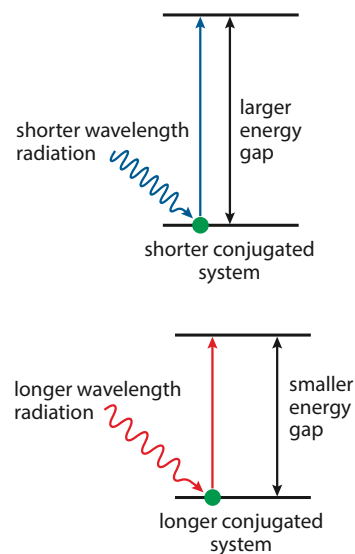

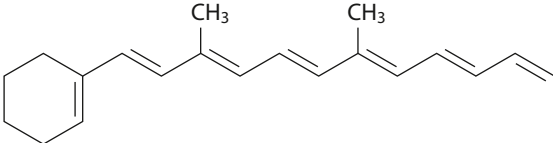
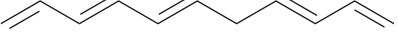
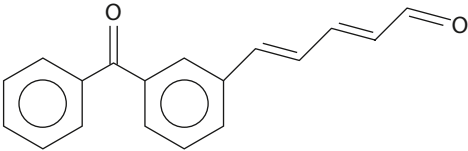


Figure C.45 As the conjugated system gets longer the energy difference between energy levels in a molecule gets smaller.

? Test yourself

35 Arrange the following molecules in order of the wavelength of ultraviolet–visible radiation that they absorb (shortest first):

	hexa-1,4-diene	
I	II	III
	buta-1,3-diene	
IV	V	VI

Learning objectives

- Understand that the electrical conductivity of a metal decreases with increasing temperature, but that of a semiconductor increases
- Understand that the conductivity of silicon can be increased by doping
- Understand how a photovoltaic cell works
- Understand how a dye-sensitised solar cell works
- Understand that the efficiency of a dye-sensitised solar cell can be increased by using TiO_2 nanoparticles
- Discuss some of the advantages of a dye-sensitised solar cell compared to a silicon-based photovoltaic cell

C8.2 Solar cells

Semiconductors

Materials can be divided into three classes according to their electrical conductivity – **metals**, **semiconductors** and **insulators**. Metals are very good conductors of electricity; insulators, to all extents and purposes, do not conduct electricity; and semiconductors have intermediate electrical conductivity.

The electrical conductivity of solids can be described in terms of band theory. The outer orbitals of the atoms making up a solid overlap to form valence and conduction bands of orbitals (Figure C.46).

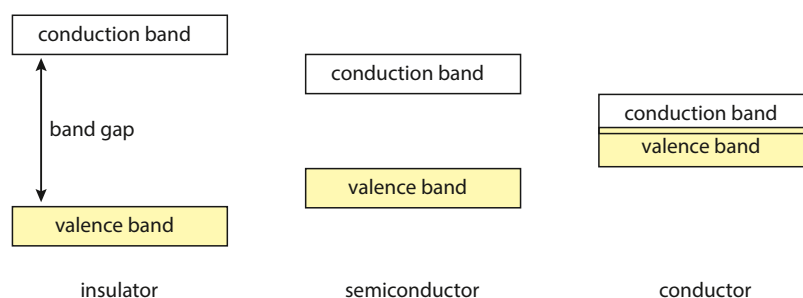


Figure C.46 Valence and conduction bands for different types of materials.

These bands consist of a very large number of orbitals that are closely spaced in energy (Figure C.47).

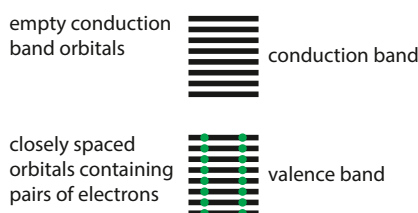


Figure C.47 Electrons in bands.

In insulators and semiconductors at absolute zero, the valence band containing the outer shell electrons of the solid is full, whereas the conduction band is empty. Because the valence band is full, there is nowhere for the electrons to move and neither insulators nor semiconductors conduct electricity at absolute zero. However, the band gap in a semiconductor such as silicon is sufficiently small that as the temperature is raised some electrons are promoted to the conduction band and become free to move. Therefore the electrical conductivity of semiconductors increases as the temperature increases. The promotion of electrons from the valence band also creates 'holes' in this band that other electrons can move into and this also contributes to the electrical conductivity (Figure C.48). A hole is regarded as a positive charge carrier – if an electron has been removed, what is left has a positive charge.

In an insulator, the energy gap between valence and conduction bands is large and at normal temperatures electrons cannot jump to the higher level. In a metal, the conduction and valence bands overlap and the metal is a good conductor of electricity because many electrons are free to move.

Metals conduct electricity well because the delocalised electrons are free to move throughout the structure. Resistance in metals arises because these electrons collide with the positive ions in the lattice (Figure C.49). As the temperature increases, the metal ions vibrate more and so there is essentially a larger cross-section for the electrons to collide with and the electrical conductivity decreases (resistance increases).

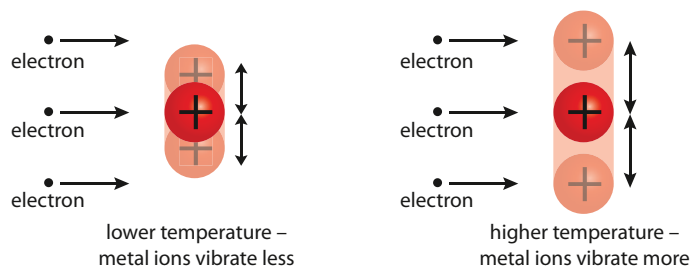


Figure C.49 Metal ions vibrate less at lower temperatures and the electrons are better able to pass through the structure.

The electrical conductivity of a semiconductor increases as temperature increases – but that of a metal decreases as temperature increases.

There is a general correlation between the electrical conductivity of metals and semiconductors and their ionisation energies. Generally, metals have the lowest ionisation energies in any period and therefore hold on to their outer electrons least strongly. This means that metals ionise fairly readily to generate free electrons. Semiconductors such as silicon have higher ionisation energies than metals and so their outer electrons are held more strongly – they will not readily ionise to form free electrons.

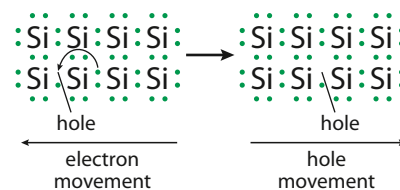


Figure C.48 The movement of a negatively charged electron in one direction is equivalent to the movement of a positively charged hole in the opposite direction.

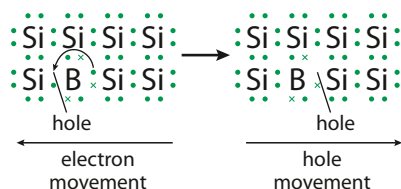


Figure C.50 The movement of holes in p-type semiconductors.

Visible light has wavelengths between 400 and 750 nm.

Doping of silicon

The electrical conductivity of silicon can be increased by incorporating atoms of group 13 or group 15 elements as impurities into the lattice – this process is called **doping**.

Inclusion of a small proportion of a group 13 element such as boron, aluminium or gallium into the silicon structure creates a p-type semiconductor (where ‘p’ stands for positive). These group 13 atoms supply only three electrons to the valence band instead of four (as supplied by Si), and therefore there is a hole in the valence band. The introduction of extra gaps in the valence band increases the electrical conductivity of the silicon. Because the main charge carriers are **positively** charged holes, this is now called a p-type semiconductor (Figure C.50).

Incorporation of a group 15 element (e.g. phosphorus, arsenic or antimony) into a silicon lattice also increases the electrical conductivity of the semiconductor, but this time by providing extra electrons. These atoms have five valence electrons, of which four are used in forming covalent bonds and go into the valence band – but the fifth is promoted to the conduction band easily. Because the current is carried by **negatively** charged electrons, the material is described as an n-type semiconductor.

The interaction of sunlight with silicon

We have seen above that electrons can be promoted from the valence band to the conduction band of a semiconductor by thermal energy. Photons of light with wavelengths shorter than about 1100 nm also have sufficient energy to cause electrons to be promoted from the valence band to the conduction band resulting in an increase in the electrical conductivity of silicon. This effect is used in photovoltaic cells (solar cells).

In photovoltaic cells, such as the solar cells found in calculators and other small electronic devices, a p-type semiconductor and an n-type semiconductor are joined together. When these two (both neutral) are joined, electrons move from the n-type to the p-type semiconductor, where they combine with some holes and leave behind positively charged ions. Holes move from the p-type to the n-type semiconductor and combine with the additional electrons, leaving behind negative charge. So, there is a build up of charge at the p–n junction that prevents any further movement of charge across the junction (Figure C.51).

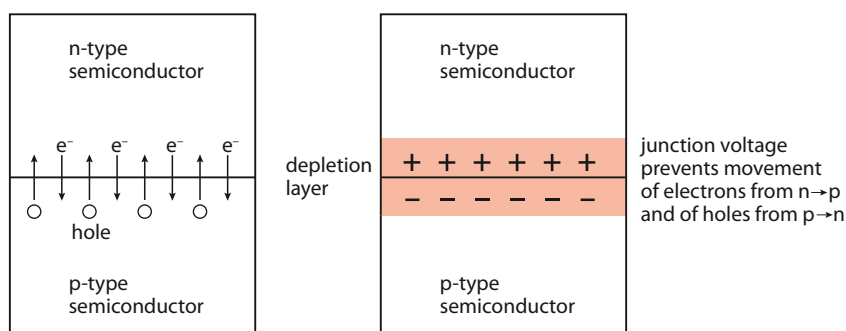


Figure C.51 The build up of charge at a p–n junction produces a layer that prevents further movement of charge across the junction.

When light hits the n-type semiconductor, electrons are promoted to the conduction band. These electrons are not able to flow from the n-type to the p-type directly, because the voltage at the p–n junction prevents this. They must therefore travel from the n-type to the p-type through the external circuit, and this can be used to power an electrical device (Figure C.52).

? Test yourself

- 36 When silicon is doped with the following atoms, would n-type or p-type semiconductors be produced?
- a Al b In c As d Sb

Dye-sensitised solar cells

During photosynthesis, light is absorbed by chlorophyll to promote electrons to higher energy levels. These electrons are then passed on via an electron transport chain (as series of redox reactions) to a low-energy electron acceptor. In the process the energy of the excited electrons is converted to chemical energy.

A dye-sensitised solar cell (DSSC or DSC) imitates the way in which a plant harnesses solar energy. The process is described in more detail below but the basic principle is that a photon of light is absorbed by a dye molecule to promote an electron to a higher energy level. The electron is passed to ('injected into') the conduction band of titanium oxide semiconductor particles and can then flow through the external circuit to the cathode. The basic idea is the conversion of light energy to electrical energy via an intermediary. Whereas in a photovoltaic cell the absorption of a photon of light causes the movement of electrons directly, in a DSSC the processes of absorption of photons and movement of electrons are separated.

Figure C.53 shows a simple diagram of a DSSC as developed originally by Michael Grätzel and co-workers.

Light passes through a glass cover with a conducting coating and causes electrons in the dye-sensitiser (D) to be promoted to a higher energy level.

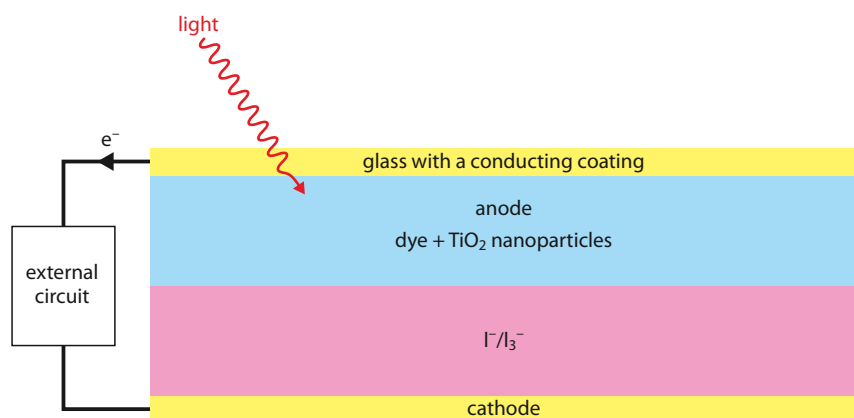
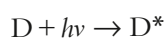


Figure C.53 A dye-sensitised solar cell.

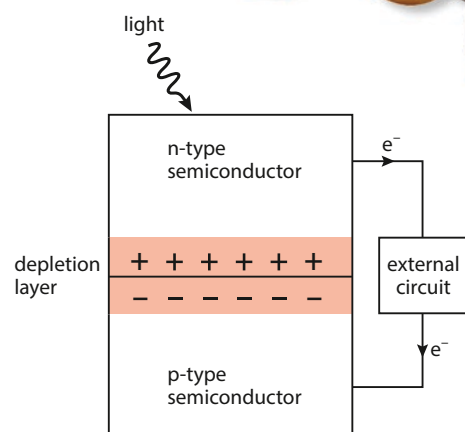
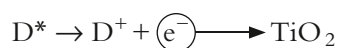


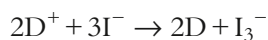
Figure C.52 How a photovoltaic cell can produce a current.

$h\nu$ represents energy from the light.

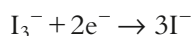
These electrons are then transferred to the conduction band of the TiO_2 nanoparticles:



The sensitiser has been oxidised and is reduced back to its original state by electrons from the I^- ions in the electrolyte:



The electrons at high energy in the conduction band of the TiO_2 flow through the external circuit to the cathode where they reduce the I_3^- ions back to I^- :



To increase the amount of light absorbed in the DSSC, porous TiO_2 nanoparticles coated with the light-absorbing dye are used – this gives an extremely large surface area. Which regions of the electromagnetic spectrum are absorbed depends on the particular dye used – for instance, a black dye gives good absorption over the entire visible spectrum.

DSSCs are still very much in the development stage but the cells produced so far have shown efficiencies similar to those of many traditional silicon-based photovoltaic cells. Several advantages over traditional silicon photovoltaic cells have been suggested:

- DSSCs are likely to be cheaper to manufacture than silicon photovoltaic cells because most of the materials involved are reasonably abundant and inexpensive.
- DSSCs are less sensitive to temperature changes than silicon photovoltaic cells – solar cells get hot in direct sunlight! The efficiency of a silicon photovoltaic cell falls with increasing temperature, whereas a DSSC is hardly affected at all.
- DSSCs can work better in low light conditions than silicon photovoltaic cells – e.g. cloudy conditions. In a silicon photovoltaic cell, electrons promoted to the conduction band by absorption of photons can simply fall back down and recombine with the holes generated at the same time, which results in no current flowing. In low light conditions the rate of recombination is likely to be significant compared to the rate of ejection of electrons. In a DSSC, no hole is generated, just an extra electron in the conduction band of the semiconductor material, and so this cannot happen.

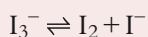
Nature of science

Solar cells in their current form were first developed in the 1950s and their principal use in the 1960s was to power satellites.

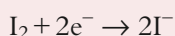
It is often important in science to have knowledge and understanding beyond one individual science and a knowledge of photosynthesis from biology provided the inspiration for scientists to develop DSSCs.

Scientists continue to make advances in the field of renewable energy – but there must also be a drive from governments and the general public to provide funding and to use the new technology. Science can only take us so far, it cannot solve all problems – scientists can develop new solar cells but they cannot make people use them. Everyone has to become more aware of their environment and the problems that we will all face in the very near future.

I_3^- is in equilibrium with I_2 and I^- in solution:



The reduction of I_3^- is thus equivalent to the reduction of molecular iodine:



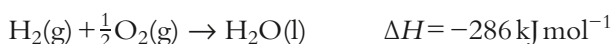


Exam-style questions

1 Renewable and non-renewable energy sources are essential in modern society.

- a Explain the difference between renewable and non-renewable energy sources. [2]
- b State **one** advantage and **one** disadvantage of using solar energy to generate electricity. [2]
- c Hydrogen–oxygen fuel cells can be used to provide power for electric vehicles. Hydrogen is stored in tanks in the vehicles and a continuous supply of hydrogen must be fed into the cell to produce electricity.

In a hydrogen–oxygen fuel cell, the reaction that occurs is equivalent to the combustion of hydrogen:



- i 10.0 mol of hydrogen were fed into a hydrogen fuel cell and 1500 kJ of electrical energy was obtained. Calculate the efficiency of the fuel cell. [2]
- ii The density of methanol (CH_3OH) at 298 K is 0.79 g cm^{-3} and the enthalpy change of combustion is -726 kJ mol^{-1} . Calculate the energy density of methanol. [3]
- iii The specific energy of hydrogen is more than six times the specific energy of methanol, but some scientists believe that methanol might be a more useful fuel for powering electric vehicles than hydrogen. Explain why this might be the case. [2]

2 It is estimated that there are over 1 billion cars in the world and the majority of these use petrol/gasoline as fuel.

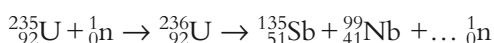
- a Explain the process by which gasoline is obtained from crude oil. [4]
- b Octane numbers are often stated on petrol pumps in petrol stations.
 - i State what is meant by the octane number of a fuel. [1]
 - ii Explain how cracking can be used to improve the octane number of fuels. [3]
- c Transport contributes a great deal to an individual's carbon footprint.
 - i Use the data below to calculate the carbon footprint per person for an aeroplane journey from London to New York. [3]

distance from London to New York	5570 km
fuel consumption of aircraft	$14.0 \text{ dm}^3 \text{ km}^{-1}$
number of passengers+crew	500
carbon dioxide produced per dm^3 of fuel consumed	2.00 kg

- ii The Ferrari 458 Italia has a fuel consumption of approximately 13.0 dm^3 per 100 km. Calculate the carbon footprint per person for two people travelling 5570 km in a Ferrari 458. Assume that the fuel has the same molecular formula as octane (C_8H_{18}) and a density of 0.700 g cm^{-3} . [5]

3 Uranium-235 is an important fuel for nuclear reactors. In a nuclear reactor the uranium-235 undergoes nuclear fission.

- a Explain what is meant by 'nuclear fission'. [1]
- b Explain in terms of nuclear binding energy why uranium-235 undergoes fission rather than fusion reactions. [2]
- c A possible reaction that uranium-235 can undergo is:

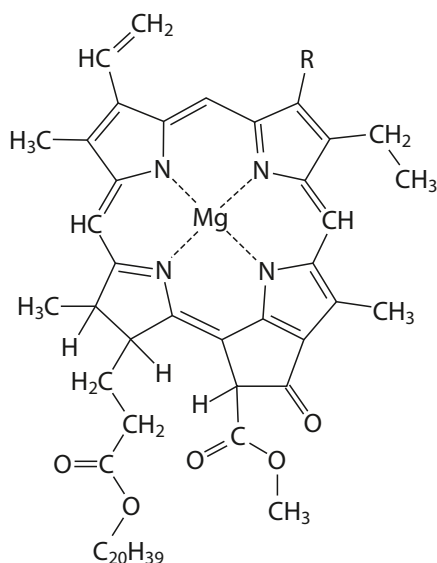


Calculate the number of neutrons produced in this reaction. [1]

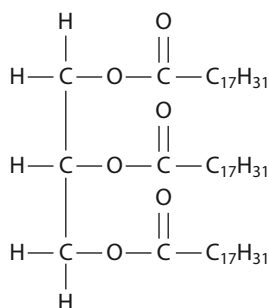
- d Antimony-135 is radioactive with a half-life of 1.7 s.
- State what is meant by 'half-life'. [1]
 - Calculate what percentage of a sample of this isotope would remain after 6.8 s. [2]
- e State **three** problems associated with generating electricity using nuclear power. [3]

4 Biofuels such as ethanol and biodiesel are becoming increasingly important as fuels for motor vehicles. The energy stored in biofuels and fuels derived from petroleum originally came from the Sun.

- a Chlorophyll is the green pigment in plants. The structure of chlorophyll is shown below. Explain what feature of the chlorophyll molecule allows it to absorb visible light. [1]



- b Light energy from the Sun is converted into chemical energy in photosynthesis. Write a balanced equation for photosynthesis. [1]
- c The structure of one of the triglycerides in sunflower oil is shown below.



This can be converted to biodiesel in a transesterification reaction with ethanol in the presence of a base catalyst.

- Write an equation for the reaction. [2]
 - Explain why the biodiesel formed in **c i** is less viscous than sunflower oil. [2]
- d Explain **one** advantage and **one** disadvantage of using liquid biofuels, such as ethanol and biodiesel, as opposed to liquid fuels derived from petroleum. [2]



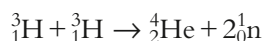
- 5 a The greenhouse effect is a natural phenomenon. It occurs as a result of the way the Earth's atmosphere interacts with radiation from and to space.
- i Explain why certain gases, such as CO_2 and CH_4 , are able to absorb infrared radiation but other gases, such as N_2 and O_2 , are not. [3]
 - ii Explain how the greenhouse effect warms the Earth. [4]
 - iii Explain why human activity may be increasing the effect of the natural greenhouse effect and state **one** consequence of this. [3]
- b Explain how both of the following statements can be true. [2]
- I Carbon dioxide is the most important greenhouse gas.
 - II Methane is a much more important greenhouse gas than carbon dioxide.
- c Explain what is meant by 'global dimming'. [2]

- HL** 6 Batteries are portable sources of electricity that have become increasingly important as we rely more and more on portable electronic devices.
- a Explain in what way a bigger battery may be better than a smaller one of the same type. [1]
- b A nickel–cadmium battery is a type of rechargeable battery often used instead of primary cells in devices such as portable radios.
- i Explain the difference between a primary cell and a rechargeable cell. [2]
 - ii Write equations for the reactions that occur at the anode and cathode when a nickel–cadmium battery is discharged and state which involves oxidation and which reduction. [3]
- c Fuels cells and rechargeable batteries have both been considered as the power source for electric vehicles. Explain why lead–acid batteries are not suitable as the main power source for an electric vehicle. [1]
- d A student carried out an experiment to measure standard electrode potentials by comparing various half-cells to a reference electrode. She obtained the value 0.32 V for the electrode potential of the $\text{Cu}^{2+}(\text{aq})|\text{Cu}(\text{s})$ half-cell. She looked up some literature values and found that the accepted value for this standard electrode potential is 0.34 V. In the conclusion to her experiment, she stated, '*my value is different to the literature value because the electrode potential varies with concentration and temperature, and therefore because the temperature in the room where I carried out the experiment was 19°C and the concentration of copper sulfate I used was 0.800 mol dm⁻³ this explains any difference between my value and the literature value.*' Comment on this statement and carry out a calculation to back up your explanation. [4]

- HL** 7 Nuclear fusion has the potential to produce vast amounts of energy.
- a State what is meant by the terms 'mass defect' and 'nuclear binding energy'. [2]
- b Explain why ^1H has a nuclear binding energy of zero. [1]
- c Tritium (^3H) is an isotope of hydrogen. Given the data in the table, calculate:
- i the mass of a tritium nucleus in unified atomic mass units [1]
 - ii the mass defect in kg [2]
 - iii the nuclear binding energy per nucleon of a tritium nucleus in J. [2]

Particle	Mass / u
proton	1.007 276
neutron	1.008 665
electron	0.000 548 6
tritium atom	3.016 05
helium atom	4.002 60

- d A possible fusion reaction is:



Calculate the energy released in this fusion reaction in kJ mol^{-1} .

[4]

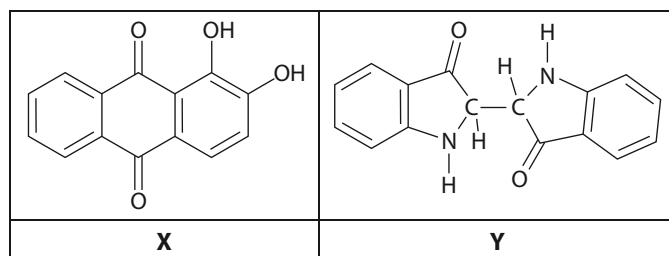
- e Tritium is radioactive with a half-life of 12.26 years. Calculate how much remains from a $100.0\mu\text{g}$ sample of tritium after 20.00 years. [3]
- f Hydrogen is a diatomic molecule and there are various combinations of the three isotopes of hydrogen that can join together to form diatomic molecules. The masses and names of the isotopes are shown in the table below:

Isotope		Mass / u
hydrogen (protium)	${}^1\text{H}$	1.007 825
deuterium	${}^2\text{H}$	2.014 102
tritium	${}^3\text{H}$	3.016 049

- i Explain the composition of the molecule that will effuse through a small hole most slowly and that which will effuse most quickly. [2]
- ii Calculate the ratio of the rates of effusion of these two gases under the same conditions. [2]

- HL** 8 a Some molecules absorb electromagnetic radiation in the UV–Vis region of the spectrum.

- i Draw the structures of penta-1,3-diene and penta-1,4-diene. [2]
- ii Predict and explain which of penta-1,3-diene and penta-1,4-diene will absorb the longer wavelength of UV radiation. [3]
- iii One of the molecules shown below is coloured. Identify the coloured compound and explain why it is coloured. [4]



- b Silicon is used in photovoltaic cells. Explain how sunlight interacts with pure silicon to enhance its electrical conductivity. [2]
- c DSSCs are an exciting new development in solar cell technology that could replace silicon-based photovoltaic cells in many applications. Explain how a DSSC, as originally designed by Grätzel, works. [4]

Option D Medicinal chemistry

D1 Pharmaceutical products and drug action

Drug therapy has come a long way since the herbal and folklore medicines of the past – the majority of drugs nowadays are synthesised in a chemistry laboratory. A large amount of research is carried out to develop specific drugs to target specific processes, in the hope that safer and more effective drugs can be developed.

The terms ‘drug’ and ‘medicine’ are often used interchangeably, but they do have slightly different definitions. A **drug** is any substance that, when applied to or introduced into a living organism, brings about a change in biological function through its chemical action. The change in biological function may be for the better – in the treatment of diseases – or for the worse – poisons that cause toxicity.

Drugs can be:

- relatively crude preparations, obtained by extracting plant or animal materials
- pure compounds isolated from natural sources
- semi-synthetic compounds, produced by chemical modification of pure natural compounds
- synthetic compounds.

The last of these is the most recent and common – most drugs are wholly synthetic.

A **medicine** is something that treats, prevents or alleviates the symptoms of disease – they have a **therapeutic** action. Medicines are usually compound preparations, which means that they contain a number of ingredients – the **active drug** itself plus non-active substances that improve the preparation in some way such as taste, consistency or administration of the drug.

A drug produces an effect on the body by interacting with a particular target molecule. This target molecule is usually a protein such as an **enzyme** or **receptor**, but may be another molecule such as DNA or a lipid in a cell membrane. When the drug binds to its target molecule, it can either stop it from functioning or stimulate it – in either case, the binding of the drug to its target produces some kind of biological effect which can either cause a beneficial (therapeutic) effect on the body or a harmful (toxic) effect.

Drug development

There are many stages involved in the drug-development process, and it can take as long as 12 years and cost hundreds of millions of dollars to bring a new drug onto the market.

Research and development of new drugs is carried out mainly by pharmaceutical companies. The decision on which disease or condition to research is based on a number of factors, probably the biggest being economic considerations – is the market big enough to give a profit? Other considerations include medical reasons (is there a medical need for

Learning objectives

- Describe the stages in the development of a drug
- Understand what is meant by therapeutic index
- Understand what is meant by therapeutic window
- Describe factors that must be considered when administering drugs
- Understand what is meant by bioavailability and some of the factors that affect it
- Understand that drug–receptor binding is dependent on the shape of the binding site

Enzymes are biochemical catalysts that catalyse nearly all the chemical reactions that occur in the body.

Receptors are proteins found on the surface of cells or inside cells that bring about a response in that cell when molecules bind to them.



Diseases of westernised countries generally generate a bigger economic return than those in developing countries – conditions such as obesity and depression are more popular targets for drug development than, for example, tropical diseases.

the new drug?) and scientific reasons (is there much known about the disease?). The ultimate goal of the research is to either find a drug that is better than existing drugs – more effective and/or with fewer side effects – or to find a drug to treat a new disease, as in the case of HIV/AIDS in the 1980s.

The first stage in the drug-development process is the identification of **lead** (rhymes with ‘seed’) **compounds**. This is done through biological testing of compounds obtained by, for example:

- isolation from natural sources
- chemical synthesis
- searching through existing ‘banks’ of compounds already synthesised.

Lead compounds have a desirable biological activity that is therapeutically relevant. They generally do not have a high amount of biological activity and are not ideal drug candidates to take forward to the clinic – for example, they may have undesirable side effects. However, they act as a starting point for **chemical modification**. A number of analogues are synthesised and tested to find more active and/or less toxic compounds which can then be developed further – this is known as **lead optimisation**.

Once a compound has been chosen for further development, the next stage is to test it for toxicity in animals (see below). Toxicity testing involves a range of different studies that look for different types of toxicities when the drug is given over different time periods. A number of drugs fail at this stage of the development process, and therefore alternative drug structures need to be identified and then developed.

Clinical trials

If a drug is found to be relatively safe in animals, it is then given to humans in clinical trials. This is the next stage of the drug-development process, and its aim is to find out if the drug is effective in humans and whether or not it is safe to use. Note that drugs may be non-toxic in animals yet toxic in humans – there may be variation in the way that different species are affected by drugs.

There are three phases of clinical trials. The first (known as **Phase I**) is carried out on a small number of healthy volunteers (usually fewer than 100) and its purpose is to find the dose range of the drug that gives a **therapeutic effect** and also to identify any **side effects**.

If the drug passes Phase I, it then enters **Phase II** clinical trials where it is tested on a small number of volunteer patients who have the disease or condition on which the drug acts. Phase II establishes whether or not the drug is effective in these patients and also identifies any side effects. If deemed safe and effective, the drug then enters Phase III.

In **Phase III** clinical trials, the drug is tested on a much larger group of volunteer patients. This phase confirms the effectiveness of the drug in the larger group and compares its activity with existing drug treatments or placebos. For example, half of the patients may be given the new drug and half given a placebo (they will not know which they have been given, and usually neither will the investigators in the study). The drug is assessed to see if it causes more of an improvement of the condition and fewer side effects in the patients to whom it has been given compared with those people given the placebo. Phase III clinical trials assess if the drug is truly

Drug trials can sometimes go disastrously wrong and in 2006 six previously healthy British men ended up seriously ill in intensive care when they took part in Phase I trials for the drug TGN1412.

effective or whether any beneficial effects seen are due to the placebo effect. Phase III trials may also identify side effects not found in previous trials because the number of patients exposed to the drug is larger.

If the drug passes Phase III clinical trials then a marketing authorisation may be obtained by the pharmaceutical company from the relevant regulatory authority; this allows the drug to enter the market to be used on patients in the wider community.

The role of chemists in the drug-development process

One of the most important roles of chemists in the development of a drug is in actually making the drug. Drugs are usually complex organic molecules and can be extremely difficult to synthesise. Initial synthesis of compounds for testing for therapeutic effects or toxicity might involve milligram amounts but once a promising compound has been selected, it is the job of the organic chemist to produce the most efficient synthetic process possible for it. A good synthesis will have as few steps as possible and produce a very good yield at each stage. The starting material(s) for the synthesis should, if possible, also be cheap and readily available. Once a drug has been synthesised it must be extracted from the reaction mixture and purified, e.g. by recrystallisation or solvent extraction. The drug must also be tested for purity to make sure that there are no unwanted compounds present. When designing a synthesis it must also be remembered that the process will have to be scaled up to make commercial amounts of the drug and that this itself can cause many problems.

Drug doses

The relationship between drug dose and physiological effect

A drug is any substance that brings about a change in biological function through its chemical action. Therefore drugs cause physiological effects on the body, and these may be therapeutic effects or side effects.


Therapeutic effect – a desirable and beneficial effect; it alleviates symptoms or treats a particular disease.

Side effect – an unintended secondary effect of the drug on the body; it is usually an undesirable effect. For example, morphine is a strong analgesic used to treat pain, but in some patients it can cause constipation, nausea and vomiting.

If a side effect is harmful to the body then it may be called a toxic effect, especially if it is caused by taking the drug in relatively large doses. For example, paracetamol (acetaminophen) can cause irreversible damage to the liver when taken in overdose.

One of the most important steps in developing a drug to treat a particular disease is determining the **dosage** of that drug – if too little is given it may not be effective; if too much is given, or it is given too often, it may be toxic.

Toxicity is sometimes assessed by determining what is known as the LD_{50} of that particular drug. LD_{50} is the dose of the drug required to kill 50% of the animals tested ('LD' stands for lethal dose). LD_{50} is expressed in units of mass per kilogram of bodyweight – if in an experimental trial,



A placebo is something that looks exactly like the real medicine but does not contain any active drug. It is made from an inert substance such as starch (if it is formulated as a tablet). Placebos are used in clinical trials on new drugs. It is found that some people who take the placebo do feel better, even though it contains only inactive ingredients. This is known as the **placebo effect**.



Measuring an LD_{50} can result in the deaths of a large number of animals – many countries have phased out this test in favour of others in which few or no animal deaths result. Another drawback with LD_{50} is that it does not give any information on long-term toxicity of a drug or toxicities that are non-lethal – for example infertility or brain damage.

The actual situation is more complicated than this and statistical analysis must be carried out on the results of tests to determine an LD_{50} value.

a dose of 500 mg kg^{-1} caused the death of 50 mice out of a sample of 100 in a certain period of time, the LD_{50} is 500 mg kg^{-1} .

A different measure of the toxicity of a drug that is also used is TD_{50} .

TD_{50} – the dose required to produce a toxic effect in 50% of the test population ('TD' stands for toxic dose).

ED_{50} – the dose required to produce a therapeutic effect in 50% of the test population ('ED' stands for effective dose).

The **therapeutic index (TI)** of a drug is the ratio of the toxic dose to the therapeutic dose – it relates the dose of a drug required to produce a desired therapeutic effect to that required to produce a toxic effect.

Exam tip

In the syllabus, *TI* for animal studies is defined solely in terms of LD_{50} .

Therapeutic index:

$$TI = \frac{LD_{50}}{ED_{50}} \quad \text{or} \quad TI = \frac{TD_{50}}{ED_{50}}$$

In humans, the definition of therapeutic index is expressed solely in terms of TD_{50} because LD_{50} studies on humans are not possible.

If a drug has a high (or wide) therapeutic index, this means that there is a large difference between the dose of the drug that causes a therapeutic effect compared with the dose that causes a toxic effect. For example, if a *TI* is 100 then TD_{50} is 100 times larger than ED_{50} , so it would require a 100-fold increase in the therapeutic dose to cause a toxic effect in 50% of the population; a high therapeutic index is therefore a desirable property of a drug. Those drugs with therapeutic indices lower than 2 are said to have a narrow therapeutic index – this type of drug must be used with caution because there is very little difference between the therapeutic dose and the toxic dose and therefore these drugs will be more likely to cause toxic effects.

Individual patients vary considerably in their response to drugs – factors such as age, sex and weight can all affect how effective (or how toxic) the drug is. Also, some conditions may require higher doses of a drug than others – for example, 75 mg of aspirin is given once daily to heart attack victims as an anticlotting agent, whereas 300–900 mg up to four times daily may be given when used as an analgesic for pain relief. It is important to know the range of doses over which a drug may be given safely – this range of doses is known as the **therapeutic window**.

Therapeutic window

A **therapeutic window** is the range of dosage between the minimum required to cause a therapeutic effect and the level which produces unacceptable toxic effects.

The therapeutic window may also be used to describe the range of concentrations of drug in the blood plasma that gives safe, effective therapy – below this range the drug would be ineffective; above it the drug would show toxic effects. At the start of therapy with a drug, blood levels of the drug are below the therapeutic level (unless it is injected directly into the bloodstream), but as the dose is repeated, blood concentration levels increase and enter the therapeutic window (Figure D.1). It is important that the dose

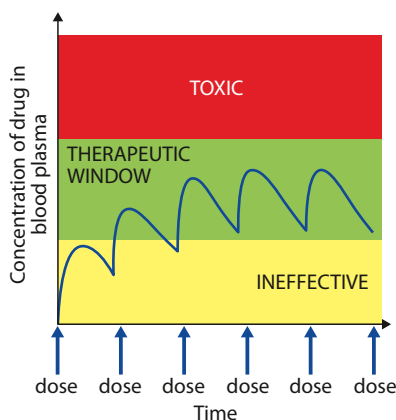


Figure D.1 Therapeutic window.



strength and frequency of dosing is such that the blood concentration of the drug is kept within the therapeutic window. This is especially important for drugs with a narrow therapeutic index, as described earlier.

Therapeutic index and therapeutic window are determined experimentally by using tests on animals and clinical trials on humans (see earlier). In animal studies, drugs are tested on healthy animals and on ones that have been infected with diseases. The effectiveness against a given disease can be determined by looking for a specific response in animals – e.g. lowering of blood pressure or the suppression of the production of a particular enzyme. Different dosages of drugs are tried on groups of animals and if, for instance, a dosage of 100 mg kg^{-1} produced a lowering of blood pressure in 50 rats out of a total sample size of 100, then this value could be taken as the ED_{50} for rats. The dosage should also be tested on other animals. LD_{50} and TD_{50} studies can be carried out in a similar way but this time the experimenters will be looking for death of the animals or indicators of toxic effects.

Tolerance

When certain drugs are given repeatedly to a patient, the intensity of the therapeutic response to a given dose may change with time, and **tolerance** to the drug may develop.

Tolerance occurs when the body becomes less responsive to the effects of a drug, and so larger and larger doses are needed to produce the same effect. This means that the patient may be at higher risk of toxic side effects.

Tolerance may develop for two possible reasons:

- repeated use of the drug stimulates increased metabolism of that drug – the body is able to prepare the drug more quickly for excretion so that lower levels remain in the body to cause an effect
- the body may adapt so that it offsets the effect of the drug – for example, by desensitising the target receptors with which the drug binds so that it is not able to produce its effect.

Addiction/dependence

When prescribing certain drugs, the possibility of dependence/addiction must be considered. Although drug addiction and dependence are usually associated with illicit drugs, addiction can also occur with therapeutic drugs. A common type of drug that people become dependent on are central nervous system depressants belonging to the class of benzodiazepines, such as diazepam (Valium®) and nitrazepam (Mogadon®).

Dependence can involve **psychological** dependence, which is the need to have the drug to feel good – the drug-taker craves the drug if deprived of it for a short time and must get further supplies in order to satisfy their need. Alternatively, it may involve **physical** dependence, in which the body cannot function without the drug – the user must keep taking the drug to avoid adverse withdrawal effects.

Dependence is also closely related to tolerance – the need to take more of the drug to produce the same effect. Benzodiazepines cause



Drugs can be beneficial but they can also have side effects. Who should

make decisions about whether a drug should be used or not? To what extent do we rely on experts to tell us what to do rather than making our own decisions? If every drug was labelled with detailed medical information concerning the benefits and adverse effects would we be better informed or just more confused? How much information do we need to make an informed choice? Can too much information be bad?

dependence and withdrawal symptoms – they have been overprescribed by doctors in the past, and some studies indicate that in many countries they are still being overprescribed. To reduce the incidence of dependence, it is advised that they should be used only in severe or distressing cases of anxiety and insomnia and not be prescribed routinely.

The administration of drugs

There are various routes by which a drug can be given to a patient. Which route is chosen is dependent on a number of factors – the chemical and physical properties of the drug, the speed at which the drug needs to act and the condition of the patient (conscious or unconscious). The five major routes of administration are oral, rectal, pulmonary, topical and by injection.

Oral

The majority of drugs are given by mouth in the form of tablets, capsules, syrups and suspensions. They pass into the stomach and intestines, and are then absorbed into the bloodstream through which they can travel to their site of action. The advantage of the oral route is that it is convenient for the patient and easy to self-administer; disadvantages are that the onset of drug action is relatively slow because the drug must first be absorbed from the gut. Also some drugs, such as insulin, are destroyed by enzymes in the gut and so cannot be given by this route.

Rectal

Drugs are incorporated into **suppositories** for administration into the rectum. They are useful if a patient is not able to take oral medication – for example, if they are unconscious or vomiting. Drugs given by this method can have either a local effect (e.g. to treat hemorrhoids) or can enter the bloodstream and have an effect on other parts of the body (e.g. morphine suppositories to treat cancer pain).

Pulmonary

Drugs are administered to the lungs in the form of gases or volatile liquids (e.g. general anesthetics) or aerosol/dry powder inhalers (e.g. to treat asthma). The lungs have a very large surface area and therefore absorption of the drug into the blood is very rapid and the drug has a fast onset of action. This route is also useful if treatment of a lung disease such as asthma is required – the drug is delivered directly to its site of action.

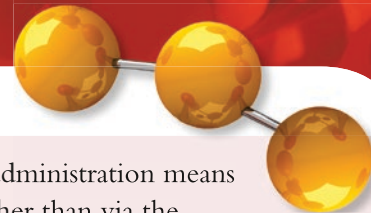
Topical

This refers to applying a drug to the skin in the form of creams, ointments or lotions. Topical administration is used primarily for local effects such as treating acne, dermatitis or skin infections, but transdermal patches (e.g. containing nicotine) may also be used and allow penetration of the drug through the skin for access to the blood circulation.

By injection

There are three main types of injection – intravenous, intramuscular or subcutaneous.

- **Intravenous** injections are the most common – they are used when a rapid therapeutic response is required because the drug is injected directly into the bloodstream.
- **Intramuscular** injections are directed into skeletal muscle, usually in the arm, thigh or buttock. Aqueous solutions of drug are rapidly



absorbed into the bloodstream, but if the drug is dissolved or suspended in oil then the drug will be released slowly from the muscle into the blood to give a sustained release of the drug over a long period.

- **Subcutaneous** injections are administered directly under the skin – absorption of the drug by the blood is slow, giving a sustained effect. Insulin is given by subcutaneous injection.

Bioavailability

The proportion of an administered drug dose that reaches the general blood circulation – and is then available to travel around the body to where it is needed (its site of action) – is known as the ‘bioavailability’ of that drug.

If a drug is given by intravenous injection, its bioavailability is 100% because all that dose is injected directly into the bloodstream. However, when a drug is given to a patient orally, not all of the dose will reach the general blood circulation.

Bioavailability is usually used in connection with drugs that are taken orally. Various factors affect the fraction of a drug dose that survives to reach the general circulation – for instance, the formulation of the tablets, their solubility, how easily it is absorbed through the intestinal wall, and the susceptibility to being broken down by enzymes in the gut and liver all affect bioavailability.

The bioavailability of a drug depends strongly on its solubility in water. Only individual molecules of a drug can pass through the wall of the intestine, therefore it is essential that a drug is soluble in water – the medium of the gastrointestinal tract. Water solubility can also affect how well a drug is transported in the blood plasma to where it is needed. Drugs that are fat-soluble will, however, pass through cell membranes (lipids) more quickly – although there are other mechanisms for drugs getting into cells. Drugs can be classified according to their solubility in water and their ability to diffuse through a cell membrane.

One of the major challenges facing chemists and pharmacologists when producing new drugs, which are often complex organic molecules, is to ensure that they are suitably soluble in water. Several factors relating to the structure of drug molecules affect solubility – the presence of polar groups (e.g. lots of OH groups) and/or functional groups that can undergo ionisation (e.g. COOH and NH₂). For instance, isoniazid (Figure D.2a), a drug used to treat tuberculosis with N–H groups that can hydrogen bond to water and other polar groups, is water-soluble but griseofulvin, an antifungal drug (Figure D.2b), is virtually insoluble in water (about 7000 times less soluble than isoniazid). Although griseofulvin has some polar groups and there will be some hydrogen bonding to water, it will not be sufficient to allow this quite large organic molecule to dissolve – most of the interactions with water around the molecule will be London forces.

It can be seen from these examples that it is not always straightforward to predict whether or not a substance will be soluble. Digoxin, a drug used to treat heart problems (Figure D.3), is virtually insoluble in water despite having a large number of OH groups – as for griseofulvin, the polar interactions are not enough to offset the non-polar ones.

‘Parenteral’ administration means any route other than via the gut – it includes injection, the pulmonary route and the topical route.

Bioavailability is quite a vague term and is defined (incorrectly) in the syllabus as the fraction of the administered drug that reaches the target part of the human body.

When a drug reaches the general circulation it will be distributed around the body – not all the drug that reaches the general circulation will reach the target site.

Exam tip

When asked to define bioavailability in the exam you should define it according to the syllabus definition: the fraction of the administered dosage that reaches the target part of the human body.

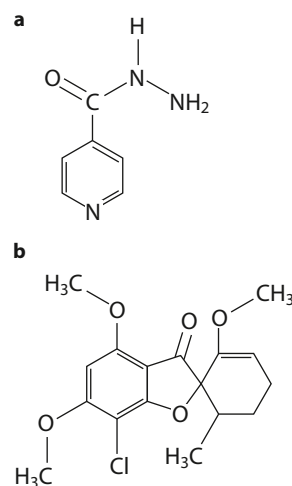


Figure D.2 a Isoniazid is water-soluble; b griseofulvin is virtually insoluble in water.

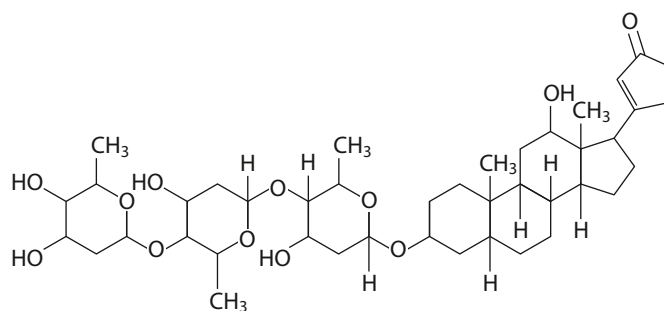


Figure D.3 Digoxin is virtually insoluble in water.

Bioavailability is also affected by the formulation of the drug – for instance, by the particle size in an orally administered drug. Just how the drug is administered is important too – as mentioned above, the bioavailability of drugs administered by intravenous injection is highest because the drug is injected directly into the blood stream.

Drug–receptor interactions

A lot of drugs act by binding to some sort of receptor in the body. These receptors are usually proteins found in cell membranes and also sometimes in the cytoplasm of cells. There has to be some sort of communication between cells in the body, and so cells have many protein molecules in their membranes that are receptors for molecular signals, for example, hormones or from nerve cells (neurotransmitters) etc. A drug can act in various ways on receptors, for example:

- it can bind to a cell-membrane protein receptor, mimicking the effect of the normal molecule that binds and cause a series of reactions in a cell – i.e. it turns a particular process in the cell on/off; in this case the drug is called a receptor agonist
- it can bind to a cell-membrane protein receptor so that the normal messenger molecule can't – it prevents a particular response from a cell; in this case the drug is called a receptor antagonist.

A drug, wherever possible, should be specific and bind to only one particular type of receptor (Figure D.4). Proteins are three-dimensional molecules with specific shapes that govern their function. The receptor binding site also has a specific shape and the ability of a drug molecule to bind to this site will depend on the shape of the drug molecule (and functional groups in the drug molecule), as well as the shape of the binding site (and specific groups in the binding site).

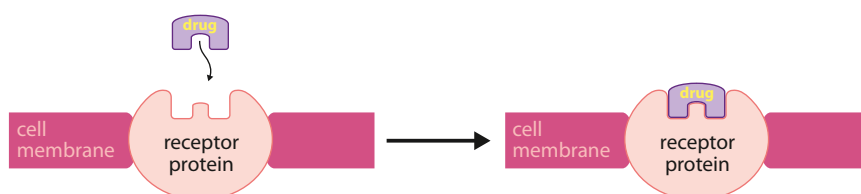


Figure D.4 The binding of a drug molecule to a receptor.



Nature of science

Scientists often have to make decisions about how much data they require to be sure about a conclusion. For instance, they must decide, based on the results of clinical trials and other evidence, whether or not a drug is safe to administer to the public. They must also sometimes consider whether the benefits outweigh the risks for a particular drug. However, the data available from clinical trials are limited and in many countries post-marketing surveillance of approved drugs, which evaluates a drug's long-term safety in the wider patient population, is in operation. In some cases, a drug that has been on the market for a number of years may be withdrawn because of serious side effects reported after widespread use.

Drugs that have been licensed and then subsequently withdrawn include terfenadine and sertindole.

D2 Aspirin and penicillin

Analgesics

Analgesics are drugs that reduce pain.

There are two main types of analgesics: **mild analgesics** and **strong analgesics**. They exert their pain-relief action in different ways. Strong analgesics will be discussed in the next section.

Mild analgesics, such as aspirin and ibuprofen, prevent the production of **prostaglandins** in the body by inhibiting an enzyme known as **cyclooxygenase** (COX), which is a key enzyme in the synthesis of prostaglandins.

Prostaglandins cause a number of physiological effects in the body, including the induction of pain, inflammation and fever.

When an injury to a tissue occurs, prostaglandins are synthesised in the damaged tissue cells and bind to receptors – this stimulates sensory nerve fibres at the site of the injury to send signals to the brain, which then interprets them as pain. They also cause dilation (widening) of the blood vessels in the damaged tissue, leading to an inflammatory response (swelling, redness, heat and pain at the site of injury) and can also stimulate the hypothalamus in the brain to cause an increase in body temperature (fever).

Mild analgesics act at the source of pain by inhibiting the production of chemical messengers that causes the sensation of pain, swelling and fever.

Aspirin

As long ago as the 5th century BCE, it was known that chewing willow bark could give pain relief. Willow bark contains a compound called salicin, which is a sugar derivative of **salicylic acid** (2-hydroxybenzoic acid) that gets converted to salicylic acid in the body. Salicylic acid (Figure D.5) is a good analgesic but causes severe irritation of the

Learning objectives

- Understand the mode of action of aspirin
- Understand why aspirin is used
- Understand that ethanol has a synergistic effect with aspirin
- Understand how aspirin is synthesised from salicylic acid
- Understand how aspirin can be purified
- Understand the characterisation of aspirin by melting point and infrared spectroscopy
- Understand how the chemical modification of aspirin can affect its bioavailability
- Understand that penicillin is an antibiotic produced by fungi
- Understand that penicillins have a β -lactam ring
- Understand how penicillins work and why the β -lactam ring is important
- Understand why modifying the side-chain in penicillin is important
- Discuss the causes of bacterial resistance to penicillin

The systematic name of aspirin is 2-ethanoyloxybenzenecarboxylic acid.

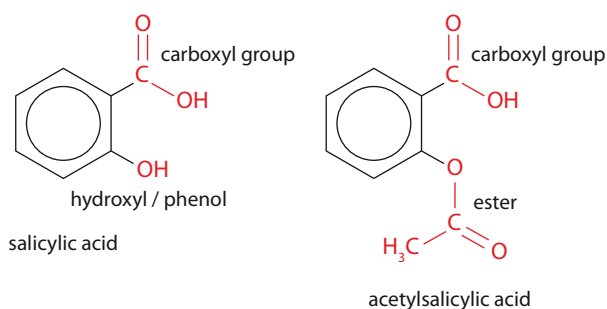


Figure D.5 The structures of salicylic acid and acetylsalicylic acid (aspirin).

stomach lining resulting in vomiting and gastric bleeding. In the 1890s, a derivative of salicylic acid, called acetylsalicylic acid (Figure D.5), began to be used medically and, over 100 years on, it is still in widespread use. Acetylsalicylic acid is the chemical name for aspirin – it is an ester of salicylic acid and is far less irritating to the stomach than salicylic acid.

Aspirin is used all over the world as an **analgesic** and **anti-inflammatory agent**. It belongs to a group of drugs known as non-steroidal anti-inflammatory drugs (NSAIDs), of which ibuprofen is also a member. It is useful in treating painful conditions such as headache, fever, and also conditions in which both pain and inflammation are present, such as arthritis.

Aspirin is also taken in low doses to help prevent recurrent heart attacks or strokes in patients who have previously suffered a heart attack or stroke – the protection is through its **anti-blood-clotting effect** – it is acting as an **anticoagulant**. Some studies have also indicated that low-dose aspirin may prevent certain cancers, in particular colorectal cancer. However, further research is needed in this area. These examples illustrate the use of aspirin as a **prophylactic** – something taken to try to prevent a disease happening in the first place.



What is pain? When we burn a finger is the pain in your finger or in your brain? When you go to the doctor, you are often asked to describe the pain – what language do we use to describe pain? Can one person ever understand another person's pain?

As we have already seen, aspirin exerts its effects through the inhibition of an enzyme called COX which plays a key role in prostaglandin synthesis. As well as mediating pain, fever and inflammation, prostaglandins also have a number of other roles in the body, one of which is maintaining the mucous layer in the stomach. Therefore, one of the side effects of taking aspirin is **gastric irritation**, both directly by the drug itself but mainly indirectly through its inhibition of prostaglandin synthesis and therefore depletion of the protective mucous layer. This can lead to peptic ulcers and possibly stomach bleeding in some patients.

Another disadvantage of using aspirin is that some people may be sensitive to it (known as **hypersensitivity**), especially those who suffer from asthma in whom aspirin can trigger an asthma attack. Another drawback of aspirin is that it is not recommended to be taken by children younger than 16 because it has been associated with **Reye's syndrome** – a potentially fatal condition that affects all organs of the body, but especially the brain and liver.



The synergistic effect of ethanol

Ethanol is an example of a drug that can increase the effects of other drugs, so care must be taken when alcoholic drinks are taken by people on certain types of medication. The increase in effect may be harmful to the body, and in some cases fatal.

Synergism can happen when two or more drugs, given at the same time, have an effect on the body that is greater than the sum of their individual effects. In other words, certain drugs can increase the effects of other drugs when given at the same time.

When alcohol is taken with aspirin there is an increased risk of **hemorrhage** (bleeding) in the stomach.

Synthesis of aspirin

Aspirin can be made from 2-hydroxybenzoic acid (salicylic acid) by warming with excess ethanoic anhydride (Figure D.6).

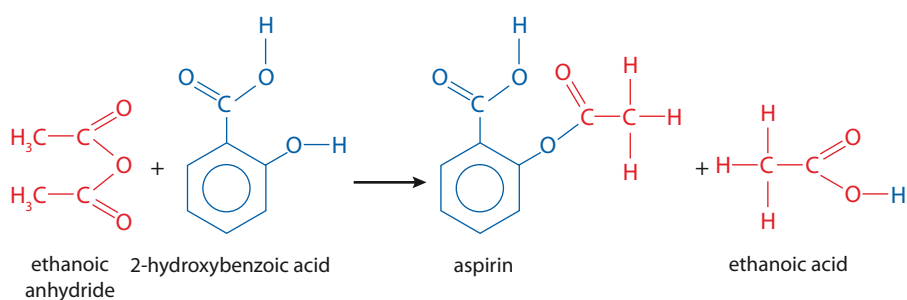


Figure D.6 Synthesis of aspirin from salicylic acid (2-hydroxybenzoic acid).

The type of reaction is **addition–elimination** (the CH_3CO group is added to aspirin and ethanoic acid is eliminated) and happens in the presence of a small amount of concentrated phosphoric (or sulfuric) acid catalyst.

Aspirin is not very soluble in water and so the addition of water to the reaction mixture causes a precipitate of aspirin to form (white solid), as well as breaking down any unreacted ethanoic anhydride to ethanoic acid. The white solid can be filtered off and washed with some cold water (to remove any soluble impurities) and left to dry (in a desiccator or warm oven) to give the crude product. The mass of the product is recorded and the yield can be worked out.

Calculation of the yield of aspirin

This is best explained using an example.

Worked example

D.1 In an experiment to synthesise aspirin, 5.60 g of salicylic acid (M_r 138.13) was reacted with 8.00 cm³ of ethanoic anhydride (density 1.08 g cm⁻³) in the presence of a concentrated phosphoric acid catalyst. 5.21 g of a white solid was obtained at the end of the reaction. Calculate:

- which reagent was in excess
- the yield of aspirin.

a The equation for the reaction is shown in Figure D.6.

$$\text{density} = \frac{\text{mass}}{\text{volume}}$$

$$\text{mass of ethanoic anhydride that reacted} = 1.08 \times 8.00 = 8.64 \text{ g}$$

$$\text{relative molecular mass of ethanoic anhydride} = 102.10$$

$$\text{number of moles of ethanoic anhydride} = \frac{8.64}{102.10} = 0.0846 \text{ mol}$$

$$\text{number of moles of salicylic acid} = \frac{5.60}{138.13} = 0.0405 \text{ mol}$$

This is a 1:1 reaction and so the ethanoic anhydride is in excess.

b To work out the yield of aspirin, we must use the number of moles of the limiting reactant, i.e. salicylic acid. From the equation, 0.0405 mol salicylic acid will produce 0.0405 mol aspirin.

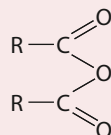
$$\text{relative molecular mass of aspirin} = 180.17$$

$$\text{theoretical yield of aspirin} = 0.0405 \times 180.17 = 7.30 \text{ g}$$

$$\begin{aligned} \text{percentage yield} &= \left(\frac{\text{actual yield}}{\text{theoretical yield}} \right) \times 100 \\ &= \left(\frac{5.21}{7.30} \right) \times 100 = 71.4\% \end{aligned}$$

Acid anhydrides

The basic structure of an acid anhydride is:



This can be regarded as being formed from two molecules of carboxylic acid with water removed (Figure D.7), although acid anhydrides are not actually made like this.

Acid anhydrides react when warmed with water to form carboxylic acids (Figure D.8). When water is added to the reaction mixture in the synthesis of aspirin, ethanoic acid is formed from excess ethanoic anhydride:

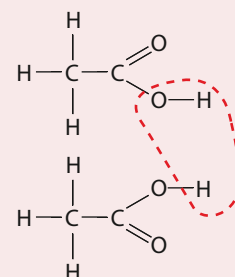


Figure D.7 Where the name 'acid anhydride' comes from.

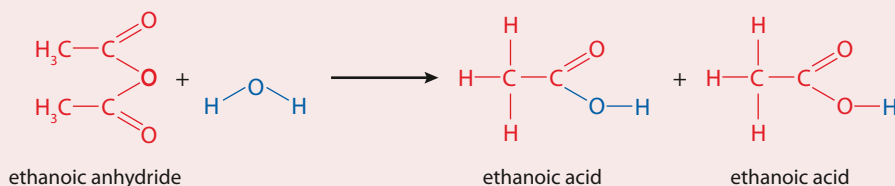


Figure D.8 Hydrolysis (breaking apart with water) of ethanoic anhydride.



Purification of aspirin

The crude sample of aspirin contains impurities and must be purified – the main impurities are unreacted salicylic acid, and possibly water if the sample is not completely dry. **Recrystallisation** can be used to purify the aspirin.

The basic principles of recrystallisation are that a solid is dissolved in a solvent in which it is soluble at raised temperatures but much less soluble at lower temperatures. Any impurities are present in much smaller amounts and so remain in solution at the lower temperature.

The procedure for recrystallisation is:

- The product is dissolved in the minimum amount of hot solvent to form a close-to-saturated solution.
- The solution is filtered while still hot to remove any insoluble impurities. Vacuum filtration is used because it is much faster – the product may start to crystallise while filtering if the solution cools too much.
- As the solution cools, the product becomes less soluble in the solvent and comes out of solution as solid crystals – less of the solid dissolves at lower temperatures. It may be necessary to cool in ice or scratch the inside of the beaker to initiate crystallisation.
- Any solid product is separated from the solvent by vacuum filtration.
- Any impurities also dissolve in the hot solvent, but because they are present in much smaller amounts they do not exceed their solubility, even at lower temperatures, and remain in solution.

Aspirin can be recrystallised from ethyl ethanoate or ethanol (usually a 95% ethanol/water mixture). Water is generally not used for recrystallisation because aspirin tends to decompose in hot water.

Characterisation of aspirin

The full characterisation of an organic compound involves determining its purity, molecular formula, physical properties, structure etc. Here we will look at how the purity of the compound can be estimated and the determination of the functional groups present in the molecule.

Determination of the purity of aspirin

How pure a sample of aspirin is can be determined by chromatography or by measuring its melting point. A pure substance will melt at a well-defined temperature but the presence of impurities lowers the melting point and causes the solid to melt over a range of temperatures. The melting point of aspirin is reported as 138–140 °C – so if a sample is tested and its melting range is found to be 125–132 °C it can be concluded that the sample is quite impure.

The infrared spectrum of aspirin

Infrared spectroscopy can be used to determine which bonds/functional groups are present in a molecule and also, by comparison with spectra in databases, to determine whether or not a particular compound has been made.

The infrared spectrum of aspirin is shown in Figure D.9.

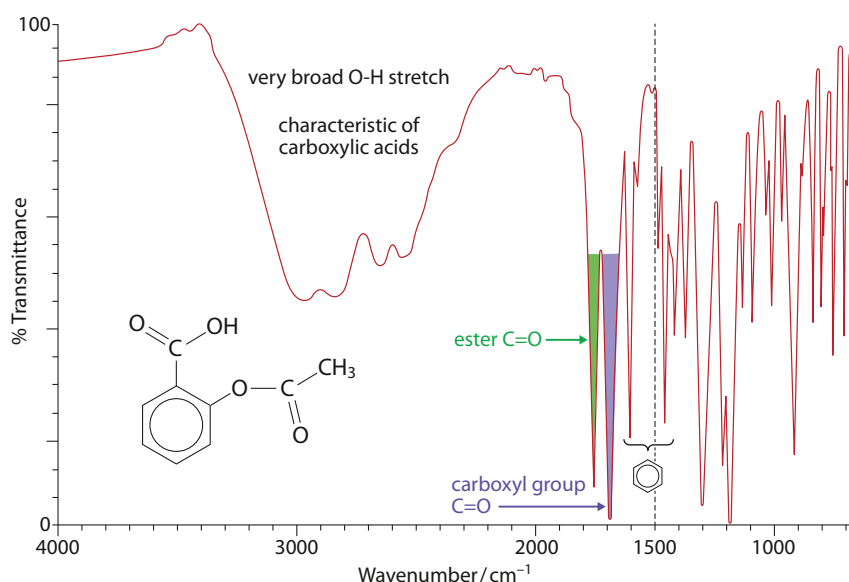


Figure D.9 The infrared spectrum of aspirin.

There are two peaks in the carbonyl ($\text{C}=\text{O}$) region due to the two different $\text{C}=\text{O}$ groups present – an ester and a carboxyl group (carboxylic acid). Consulting of more advanced tables of infrared data allows us to assign each peak as shown. The peaks at 1600 cm^{-1} and just below 1500 cm^{-1} are due to the vibrations of $\text{C}=\text{C}$ bonds in the benzene ring.

If the infrared spectrum of aspirin is compared with that of salicylic acid (Figure D.10), the spectra are very similar but the $\text{C}=\text{O}$ stretch from the ester at just above 1700 cm^{-1} is missing.

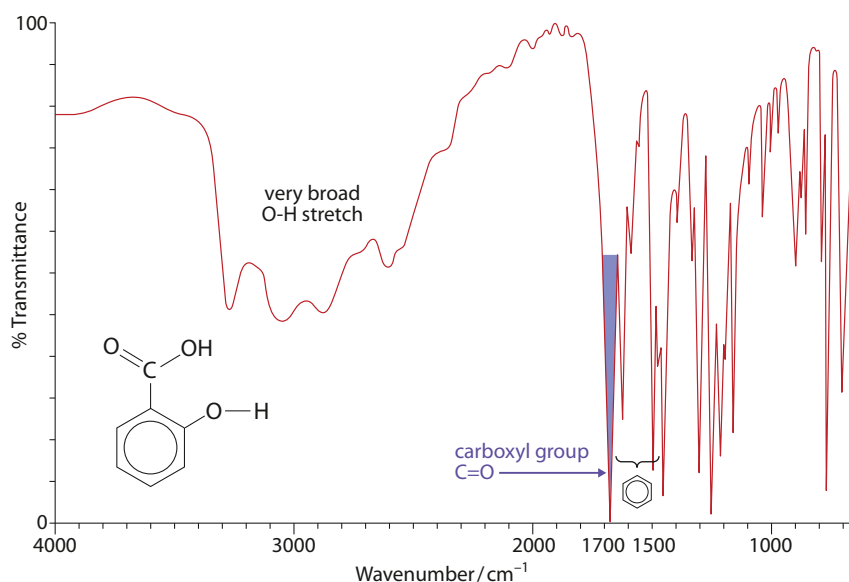


Figure D.10 The infrared spectrum of salicylic acid.

Solubility of aspirin and other drugs

Aspirin is administered orally and therefore must first be absorbed from the gastrointestinal tract before reaching the blood circulation to be distributed to the various body tissues. For a drug to enter the blood circulation after oral administration, it must first dissolve in the aqueous environment of the intestines before it can be absorbed across the lipid membranes of the intestinal wall. If the rate at which the drug dissolves is

slower than the rate at which it gets absorbed, this can affect the amount of drug that gets absorbed – and hence its bioavailability. Once in the bloodstream, the drug has to travel through the aqueous blood plasma and be distributed through the body to reach its site of action.

One way to increase the aqueous solubility of an acidic or basic drug is to make the ionic salt of the drug. Aspirin is an example of an acidic drug – it has a **carboxyl (carboxylic acid)** group that can be reacted with a strong alkali to form a salt. This converts the acid group into the **anion** (COO^-). The most common salts of acidic drugs are their sodium salts, and the formation of the sodium salt of aspirin is shown in Figure D.11. The sodium salt of aspirin is more water-soluble than aspirin and so is absorbed more rapidly into the bloodstream, increasing its bioavailability.

Many drugs contain an amine (amino) group, such as the opioid analgesics, amphetamines and some antidepressants. Because the amine group is basic, these drugs can be converted into salts by reacting the amine group with a strong acid, such as hydrochloric acid, to produce the cation. The most common type of salt for basic drugs is the chloride salt, formed by reacting the amine group with hydrochloric acid. The formation of fluoxetine hydrochloride (Prozac[®]) is shown as an example in Figure D.12.

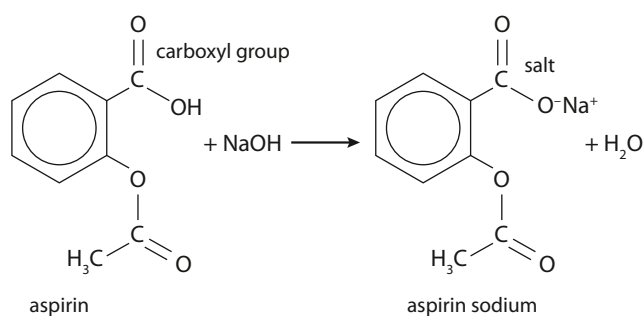


Figure D.11 Conversion of aspirin into aspirin sodium.

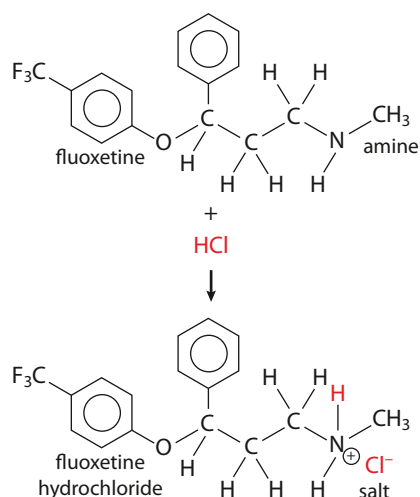
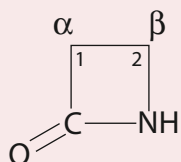


Figure D.12 Conversion of fluoxetine into fluoxetine HCl.

Cyclic amides are named using Greek letters to indicate the size of the ring. So a γ -lactam has a five-membered ring and a δ -lactam has a six-membered ring. The Greek letter refers to which carbon, going round the ring from the C=O group, the N atom is joined to – for example, the second or β -carbon in a 4-membered ring.



Penicillin

Antibacterial drugs are some of the most frequently prescribed medicines. These drugs are toxic to bacteria while being relatively safe to the patients who take them. They achieve this by acting on sites in the bacterial cells that are either different from those in our cells or that do not exist in our cells at all.

There are many different types of antibacterial drugs (commonly called antibiotics), but the most commonly prescribed are the **penicillins**. They were discovered by chance in 1928 by a Scottish physician and microbiologist called Alexander Fleming. Penicillins are produced by some fungi of the *Penicillium* strain, such as *Penicillium chrysogenum*. One of the most important natural penicillins is benzylpenicillin (penicillin G) and this is manufactured by fermentation of a mixture of corn-steep liquor (a byproduct of corn-starch manufacture), sugars, minerals and phenylethanoic acid using a penicillin fungus in a carefully controlled environment.

Penicillin has a bicyclic structure (Figure D.13) containing a **β -lactam** ring (a cyclic amide that is part of a four-membered ring). This β -lactam ring is essential for the antibacterial activity of penicillin; if the ring is broken in any way, such as by acid or bacterial enzymes (see below), the penicillin is no longer active.

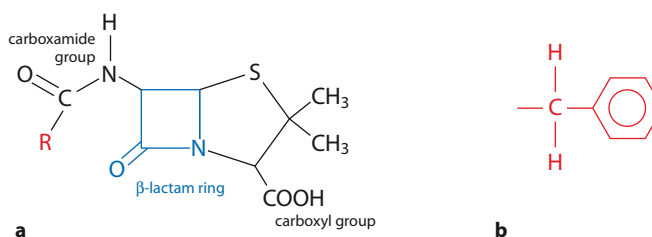


Figure D.13 All penicillins have the same basic bicyclic structure, but different penicillins have different side-chains. **a** The general structure of penicillins; **b** the side-chain for benzylpenicillin (penicillin G).

Action of penicillin on bacterial cell walls

Bacterial cells differ from our own cells in that they contain a cell wall which contains a polymer made up of sugar chains cross-linked with peptides (short stretches of amino acids). This polymer has a mesh-like structure and gives strength to the cell wall, allowing the bacteria to withstand high osmotic pressures. Penicillin acts by irreversibly inhibiting an enzyme (transpeptidase) involved in the cross-linking of this polymer, resulting in a weakened cell wall and causing the bacterial cell to burst due to the high osmotic pressure caused by water from the surroundings entering the bacterial cell. Penicillin is not the only antibacterial that works by inhibiting cell-wall synthesis – cephalosporins and carbapenems work in a similar way.

The β -lactam ring is essential to the mode of action of penicillin (Figure D.14). An OH group on the side-chain of an amino acid (serine) in the transpeptidase-enzyme active site reacts with the β -lactam ring of the penicillin instead of its normal substrate. A covalent bond is formed between the enzyme and penicillin as the β -lactam ring opens – the complex formed prevents any substrate molecules entering the active site and reacting, therefore the enzyme is deactivated.

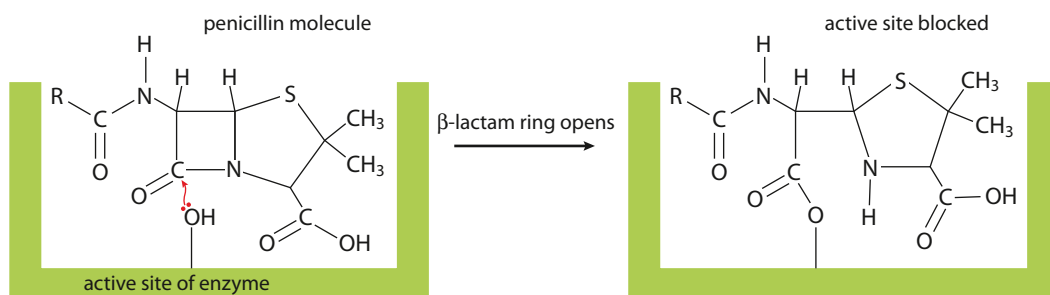


Figure D.14 The mode of action of penicillin.

The first penicillin to be isolated and purified was penicillin G (benzylpenicillin) (Figure D.13). However, this penicillin has a number of disadvantages, one of which is that it is easily broken down by stomach acid and must be given by injection. Scientists have overcome this problem by making derivatives of penicillin G that have modified side-chains (R in the general penicillin structure in Figure D.13a) that can resist stomach acid and be given by the oral route.

Bacterial resistance

The widespread use of penicillins has resulted in the development of bacteria that have become resistant to their antibacterial effects – this is known as **bacterial resistance** and arises because of mutations in the DNA of bacteria to aid their survival. Some strains of bacteria have developed ways of counteracting the effects of certain penicillins by producing an enzyme known as **penicillinase** (a β -lactamase), which opens the β -lactam ring of the penicillin, rendering it inactive. Penicillin G is an example of a penicillin that is inactivated by penicillinase. However, scientists have now developed penicillins that are less sensitive to the effects of this enzyme by modifying the side-chain in the penicillin structure (Figure D.15).

Bacterial resistance has developed not just for penicillins, but for most other types of antibacterials too. Some bacteria are resistant to more than one type, making them extremely difficult to kill, so it is important to carry out research into the discovery and development of new antibacterial agents.

It is extremely important that antibacterials are taken according to a doctor's instructions (called **patient compliance**) and that the whole course of treatment is taken. Otherwise failure to kill all the bacteria in the infection can lead to development of resistance in those bacteria that survive.

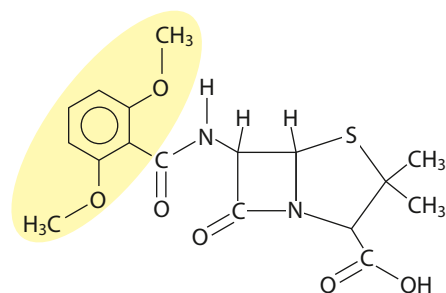


Figure D.15 Methicillin has a different R group and is resistant to penicillinase enzymes. However, some strains of bacteria have become resistant to methicillin. MRSA, one of the so-called 'superbugs', stands for 'methicillin-resistant *Staphylococcus aureus*'.

Penicillin G can be used to treat diseases caused by bacteria that do not produce penicillinase, such as meningitis and gonorrhea.

Modifying the side-chain in penicillins makes them more resistant to the penicillinase enzyme.

Such widespread bacterial resistance is also due to the extensive use of antibacterials, both for human use and for animals. Overprescribing of antibacterials for minor infections has increased the exposure of bacteria to the antibacterial agents and has increased the number of resistant bacteria. Antibacterials are also used extensively in animal feeds to lower the occurrence of infections in livestock. These antibacterials are given to healthy animals and can result in the development of resistant bacteria that can be passed on to humans via meat and dairy products.

Bacterial resistance is a widespread problem – it has developed because of the innate ability of bacteria to mutate DNA in order to survive in hostile environments, as well as the overuse and misuse of antibacterials. Improving the way that antibiotics are prescribed and taken by humans or used for livestock is essential if the development and spread of resistant bacteria is to be controlled.

Nature of science

Many scientific discoveries come about following a systematic approach to research but some discoveries can be the result of a chance set of conditions and serendipity. The discovery of penicillin was one such situation but the genius of the scientist who discovered penicillin was in recognising that he was seeing something different – not everyone would have made the connections required.

Learning objectives

- Understand what is meant by an opiate
- Understand the mode of action of strong analgesics such as morphine and codeine
- Compare the structures of morphine, codeine and diamorphine
- Explain why diamorphine is more potent than morphine
- Understand how diamorphine and codeine can be synthesised from morphine
- Explain the advantages and disadvantages of using opiates

The term ‘narcotic’ can be used in different ways. It is used here to describe analgesic drugs derived from opium, but nowadays it is often used in everyday language to indicate any illicit/strictly controlled drug.

D3 Opiates

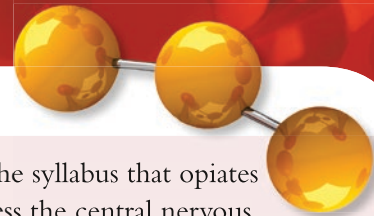
Strong analgesics

Whereas mild analgesics, such as aspirin, are used for relatively mild pain, such as headache or toothache, opiate/opioid analgesics are strong analgesics used for moderate to severe pain, such as in terminally ill patients. Mild analgesics may be combined with strong analgesics in some preparations – for example, paracetamol and codeine are often used together.

Opiates

Opiates are natural narcotic (sleep-inducing) analgesics derived from the opium poppy.

Opiates are derived from the juice of the unripe seed pods of the poppy *Papaver somniferum*. This juice is known as **opium** (the Greek word for ‘juice’) and contains a mixture of approximately 25 different nitrogen-containing compounds (known as **alkaloids**), the most important of which is **morphine**. Morphine was first isolated in 1803 and is chiefly responsible for the biological effects of opium – it accounts for approximately 10% of the opium mixture. **Codeine**, a milder analgesic than morphine, is also found naturally in opium, although in smaller proportions.



Strong analgesics work by temporarily binding to opioid receptors in the brain, which block the transmission of pain signals in the brain.

Morphine and codeine are strong analgesics, which act by temporarily binding to **opioid receptors** in the brain. This blocks the transmission of pain signals in the brain and increases the pain perception threshold – even though pain in the affected tissue is still occurring and being transmitted via the peripheral nervous system, the patient is not as aware of it. Also, opioids increase the tolerance to pain, which means that even if pain is felt by the patient they are more able to tolerate it.

Opiates cause a number of effects on the body through binding to opioid receptors. These include analgesia, sedation, a feeling of well-being and suppression of the cough reflex. They are used medically for pain relief and the treatment of coughs and diarrhea.

Opioid receptors in the brain are essential for the action of opiates such as morphine. These opioid receptors are proteins and there are various types in the brain. However, the opioid receptor that causes the greatest analgesic effect when opiates bind to it is also the one responsible for the greatest side effects, such as euphoria, addiction etc.

Both the medicinal effects of opiates and their addictive properties are caused by binding to the same opioid receptors in the brain.

It states on the syllabus that opiates do not depress the central nervous system. However, the brain is part of the central nervous system and opiates are CNS depressants.

Exam tip

When asked about the mode of operation of strong analgesics in the examination you should use the definition given on the syllabus: 'strong analgesics work by temporarily bonding to receptor sites in the brain, preventing the transmission of pain impulses without depressing the central nervous system'.

Structures of morphine and its derivatives

The chemical structures of codeine, morphine and diamorphine are shown in Figure D.16. As can be seen, they are very similar in structure – all have a tertiary amine group and benzene ring, which are essential for analgesic activity.

The only difference between codeine and morphine is a **methoxyl** ($-\text{OCH}_3$) group (ether functional group) on the benzene ring in codeine instead of a **hydroxyl** ($-\text{OH}$) group (an OH group attached directly to a benzene ring gives rise to a phenol) in morphine. When codeine enters the body, some of it is acted on by enzymes, which remove the methyl group to give a hydroxyl group; thus codeine is converted to morphine.

A tertiary amine has N joined to three C atoms (three alkyl groups).

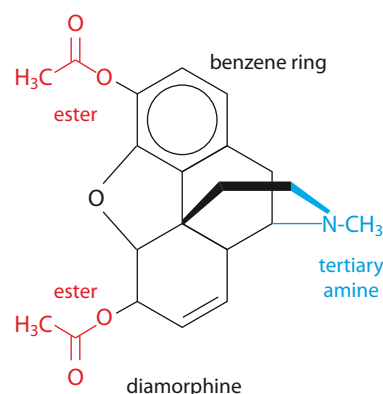
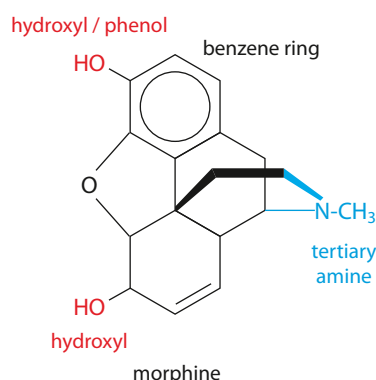
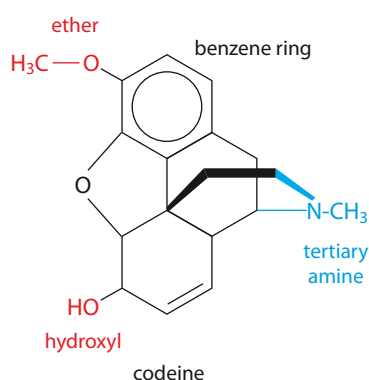


Figure D.16 Structures of codeine, morphine and diamorphine.

Diamorphine is not a naturally occurring substance derived from poppies – it is made from a product derived from opium, so it does not fit the definition of an opiate given above. The definition of an opiate is, however, usually extended to include semi-synthetic morphine-like substances derived from morphine. In some definitions, diamorphine is described rather as an **opioid**, which is a wider class of compounds exhibiting morphine-like effects on the body – opiates are opioids, but not all opioids are opiates. The terms ‘opioid’ and ‘opiate’ are often used interchangeably.

It is this conversion to morphine that accounts for the therapeutic properties of codeine, which suggests that the phenol group is also essential for the analgesic activity of opiates.

Diamorphine (heroin) (Figure D.16) is a **semi-synthetic** morphine derivative. The difference between the structures is that diamorphine contains two ester (CH_3COO) groups, whereas morphine contains two OH groups.

Diamorphine is a more potent analgesic than morphine because it is better able to cross the blood–brain barrier.

Diamorphine is more lipid-soluble than morphine because of the replacement of the OH groups (which can take part in hydrogen bonding) by the ester groups (which cannot) and therefore is able to cross the **blood–brain barrier** and enter the brain more easily. The blood–brain barrier is essentially a lipid barrier that prevents the entry of potentially toxic substances from the capillaries into the brain – it allows small, lipid-soluble molecules across and hinders large, polar molecules. Once diamorphine has entered the brain, it is hydrolysed by enzymes to the monoester (only one ester group) and to morphine; these bind to opioid receptors and produce an analgesic effect.

Synthesis of derivatives of morphine

Diamorphine

Diamorphine is synthesised from morphine by heating it with ethanoic anhydride (Figure D.17). This converts the two hydroxyl groups in morphine to ester groups. The type of reaction that occurs is addition–elimination (as in the synthesis of aspirin on page 11) – it could also be called **esterification**. $\text{CH}_3\text{COO}-$ is the ethanoate group and so two ethanoate esters are formed.

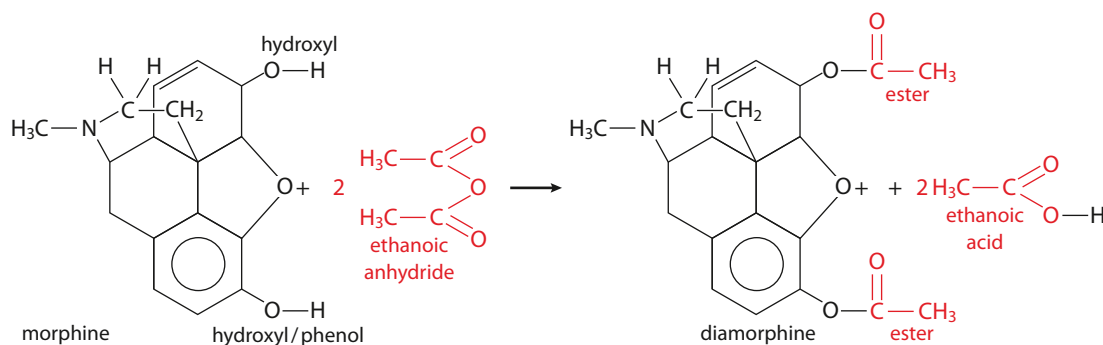


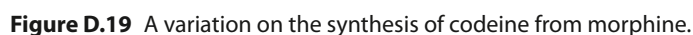
Figure D.17 Synthesis of diamorphine from morphine.

Codeine synthesis

Codeine can also be synthesised from morphine (Figure D.18). In the original process, morphine was reacted with iodomethane (the methylating agent) in the presence of a base. Phenols are slightly acidic and so the presence of a strong base converts the OH of the phenol to O^- . The reaction is nucleophilic substitution, with the O^- attacking the δ^+ carbon atom of the CH_3I .



The synthesis is more usually carried out nowadays using a more complicated methylating agent – a salt of $\text{C}_6\text{H}_5\text{N}(\text{CH}_3)$ such as $\text{C}_6\text{H}_5\text{N}(\text{CH}_3)^+(\text{C}_2\text{H}_5\text{O}^-)$ – Figure **D.19**.



Opiates such as morphine and diamorphine are used medically for the relief of severe pain – they are especially effective in visceral pain (pain in internal organs, such as the liver and lungs). They are commonly used to relieve the pain associated with cancer in terminally ill patients. Morphine may also be used for the short-term control of diarrhea due to its constipating effect, and to control distressing coughing by lung cancer patients, due to its cough-suppressant effect. Milder opiates such as codeine are used to relieve moderate pain. Codeine is also used as a cough suppressant for dry coughs and as an antidiarrhea drug.

Opiate analgesics have a number of side effects associated with their use – in the short term they can cause nausea and vomiting, constipation, respiratory depression (slowed or shallow breathing), drowsiness and euphoria; in the long term they cause **dependence** and **tolerance**, chronic constipation and decrease in sex drive.

There are two types of dependence:

- **psychological dependence**, in which the drug-taker craves the drug if deprived of it for a short time and must get further supplies in order to satisfy their need
- **physical dependence**, in which the body cannot function without the drug and deprivation results in withdrawal symptoms.

Illicit drug users suffer both physical and psychological dependence, whereas patients taking opioids for medical reasons generally suffer only physical dependence. Tolerance occurs in both types of user, requiring higher doses to be taken to cause the same effect (therapeutic or euphoric).

Learning objectives

- Understand that antacids can be used to reduce the amount of excess acid in the stomach
- Understand that the action of antacids is non-specific
- Write equations for neutralisation reactions involving different antacids
- Understand how ranitidine (Zantac[®]) works
- Understand how omeprazole (Prilosec[®]) and esomeprazole (Nexium[®]) work
- Understand what is meant by an active metabolite
- Solve problems involving buffer solutions

Abuse of opiates

Opiates have been taken for non-medical reasons for centuries. As well as dulling pain, they cause a pleasant, dreamy and relaxed state known as **euphoria**, with heroin also causing a feeling of warmth and thrill when injected intravenously. Because heroin is lipophilic, it enters the brain quickly and so causes a 'euphoric rush'. However, dependence and tolerance develop quickly, and the user soon starts to need larger and larger doses to retain this 'rush'. If the user is denied the drug withdrawal symptoms occur, including anxiety, cold sweats, vomiting and jerking of the legs. Treating opiate dependence is difficult – it may involve a gradual reduction of the dose of the drug and the administration of a substitute called **methadone** which also binds to opioid receptors but has a prolonged action and reduces the craving and prevents withdrawal symptoms.

Opiate dependence is a worldwide problem and is associated with a significant amount of crime. Users may find that they can no longer afford to pay for the increasing doses needed and so resort to criminal activity to pay for their drugs. Users who inject heroin intravenously are also at increased risk of infection from hepatitis or HIV/AIDS by sharing needles.

Nature of science

Scientific knowledge is continually developing. Although opium has been known and used for thousands of years it is only now that our knowledge of biochemistry has developed sufficiently for us to understand its mode of action on the molecular level.

D4 pH regulation of the stomach

Normally the pH in the stomach is between 1 and 2, owing to the production of hydrochloric acid by the millions of gastric glands that line the stomach. The stomach is maintained at such a low pH for two main reasons:

- the acidic environment is not tolerated by the majority of microorganisms (e.g. bacteria) that may enter the digestive system with food – the low pH plays a role in the body's natural defence against disease-causing microorganisms
- the digestive enzymes in the stomach (e.g. pepsin, which breaks down proteins) require a low pH for optimum catalytic activity.

A layer of mucus lines the stomach, and protects the stomach wall from damage by the acid. However, irritation to the stomach lining can occur by the production of excess acid – for example, caused by drinking too much alcohol, eating large (especially fatty) meals, smoking or stress. Certain drugs can irritate the stomach lining directly, whereas drugs such as aspirin can lower the production of mucus in the stomach making the stomach lining more susceptible to acid attack. This can result in the following:

- **indigestion** – irritation of the stomach lining caused by excess acid producing pain or discomfort in the upper abdomen and/or nausea
- **heartburn** (acid reflux) – acid from the stomach rising up into the esophagus causing a burning sensation
- **peptic ulcer** – erosion of part of the gut lining, caused by acid



penetrating the mucous layer. This can be a serious condition if left untreated because internal bleeding can occur. Aspirin and other related anti-inflammatory drugs can cause ulcers in some patients.

Antacids are used to treat these conditions. They are weakly basic compounds that neutralise acids, relieving the pain, discomfort or burning sensation and allowing repair of the mucous layer. In the case of peptic ulcers, neutralisation of the acid prevents further erosion of the gut lining allowing ulcers to heal.

The most commonly used antacids are metal hydroxides, carbonates and hydrogencarbonates (bicarbonates):

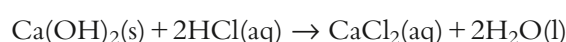
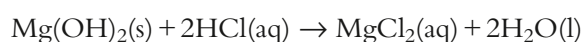
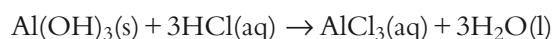
- magnesium hydroxide
- aluminium hydroxide
- calcium carbonate
- sodium hydrogencarbonate (also called sodium bicarbonate).

Some antacid preparations contain mixtures of two different antacids, such as magnesium compounds and aluminium compounds (usually magnesium and aluminium hydroxides). The rationale for using these two different antacids is that magnesium salts are faster acting and so work quickly to neutralise the acid, but aluminium salts have a slower and more prolonged effect, so the time interval between doses is increased. Also, magnesium salts in repeated doses can cause a laxative effect, but this is offset by aluminium salts which can induce constipation.

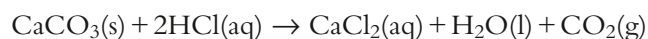
Unlike the other drugs that have been discussed above,

**antacids are non-specific and do not bind to protein receptors.
They work by simply neutralising excess stomach acid.**

The neutralising reactions for hydroxides are:



Metal carbonates and hydrogencarbonates also react with the acid to give a salt along with water and carbon dioxide:



The term **dyspepsia** is often used interchangeably with indigestion but it is defined more generally as pain or discomfort in the upper abdomen.

Exam tip

Calcium hydroxide and sodium carbonate are also mentioned on the syllabus as antacids but these are not generally given in antacid preparations – presumably because they are also irritants.

Because carbon dioxide can cause bloatedness and flatulence, **antifoaming agents** may sometimes be included in a preparation – for example, activated **dimeticone (dimethicone)**, which relieves flatulence.

Alginates may also be present in some antacid preparations. These form a 'raft' that floats on top of the stomach contents reducing reflux into the esophagus, which causes heartburn.

Worked example

D.2 Compare the volume of stomach acid (hydrochloric acid) of pH 1.50 that is neutralised by taking one indigestion tablet containing 1.00 g of calcium carbonate with one containing 1.00 g of sodium hydrogencarbonate.

A pH of 1.50 corresponds to a concentration of $\text{H}^+(\text{aq})$ of $10^{-1.50} = 0.0316 \text{ mol dm}^{-3}$

Because HCl is a strong acid it completely dissociates and the concentration of $\text{H}^+(\text{aq})$ is equal to the original concentration of the acid.

The equation for the reaction with calcium carbonate is:



$$1.00 \text{ g of CaCO}_3 \text{ is } \frac{1.00}{100.09} = 9.99 \times 10^{-3} \text{ mol}$$

$9.99 \times 10^{-3} \text{ mol CaCO}_3$ reacts with $2 \times 9.99 \times 10^{-3}$ moles of HCl

i.e. 0.0200 mol hydrochloric acid

$$\text{volume of hydrochloric acid} = \frac{0.0200}{0.0316} = 0.632 \text{ dm}^3 \text{ or } 632 \text{ cm}^3$$

$$\text{volume} = \frac{\text{number of moles}}{\text{concentration}}$$

The equation for the reaction with sodium hydrogencarbonate is:



$$1.00 \text{ g of NaHCO}_3 \text{ is } \frac{1.00}{84.01} = 0.0119 \text{ mol}$$

0.0119 mol NaHCO_3 react with 0.0119 mol hydrochloric acid

$$\text{Volume of hydrochloric acid} = \frac{0.0119}{0.0316} = 0.377 \text{ dm}^3 \text{ or } 377 \text{ cm}^3$$

1.00 g of calcium carbonate therefore reacts with significantly more hydrochloric acid. This is because the molar masses are fairly similar and each mole of calcium carbonate reacts with twice as many moles of hydrochloric acid as sodium hydrogencarbonate does.

? Test yourself

1 Work out the volume of hydrochloric acid of pH 2.00 that reacts with:

- a** 1.00 g of aluminium hydroxide
- b** 1.00 g of magnesium hydroxide

Treatment of peptic ulcers

Stomach acid is produced by parietal cells, which are cells in the lining of the stomach. The treatment of peptic ulcers involves regulating the acid levels in the stomach. There are two main approaches to this – stopping the production of the acid and preventing the release of the acid into the stomach.

Ranitidine

Ranitidine or Zantac[®] (Figure D.20) is a drug that inhibits the production of acid. It does this by binding to a receptor protein (histamine H₂-receptor) in the membrane of the parietal cells, which stops the normal chemical messenger (histamine) from binding to turn on the chain of events for producing acid. Ranitidine therefore prevents the production of stomach acid.

‘H₂’, in this instance, has nothing to do with hydrogen gas.

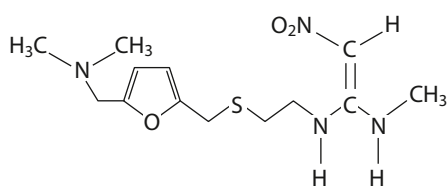


Figure D.20 The structure of ranitidine.

Ranitidine can be described as an **H₂-receptor antagonist** because when it binds to an H₂-receptor it does not cause activation of the receptor, but rather stops the naturally occurring molecule that does cause activation (the agonist) from binding.

Omeprazole and esomeprazole

Omeprazole (Losec[®], Prilosec[®]) and esomeprazole (Nexium[®]) (Figure D.21) are **proton pump inhibitors** and work by preventing the release of acid from the parietal cells into the stomach. Protons are released from the parietal cell by the action of a proton pump. This is a protein complex that moves protons through cell membranes – being charged, protons cannot diffuse normally through a cell membrane made of mainly non-polar lipid molecules.

These drugs are weak bases but are mainly in the un-ionised form at the pH of blood plasma. They are also mostly non-polar and therefore lipid-soluble so they can pass through the cell membrane of the parietal

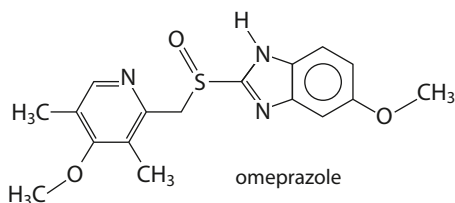


Figure D.21 The structure of omeprazole. Esomeprazole is a stereoisomer of this – the atoms are joined together in the same order but arranged differently in space.



Esomeprazole (Nexium[®]) is one of the biggest-selling prescription drugs in the world, and at times has been the biggest-selling prescription drug in the US.

cells. Inside the parietal cells the medium is much more acidic and the basic molecules get protonated. Protonation starts a series of reactions that changes the structure of the drug molecule into one that can bind irreversibly to the proton pump (Figure D.22) and so stop it from carrying out its function. The drugs are effective for an extended period of time – until the cell is able to make new proton pumps.

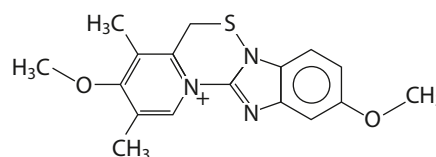


Figure D.22 The active form of omeprazole.

Active metabolites

We have already seen examples of drugs that are converted into a different form in the body – the form that causes the desired action of the drug. So, for instance:

- **codeine** is converted into morphine in the body and it is the morphine that binds much more strongly to the opioid receptors than codeine, producing an analgesic effect
- **omeprazole/esomeprazole** are converted into different forms that are able to bind to proton pumps
- **aspirin** is converted into the active form – salicylic acid. Salicylic acid cannot be taken orally because it causes severe irritation of the stomach lining, resulting in vomiting and gastric bleeding. Therefore it is taken in ester form; this causes much less gastric irritation but is converted back into the active analgesic in the body.

Active metabolites are the active forms of drugs after they have been processed in the body.

There are many reasons for making a drug in a different form to that of the active metabolite and these include:

- to avoid side effects – e.g. aspirin
- to allow the drug to pass through cell membranes – the active form of omeprazole is charged and would not pass through the cell membrane into the parietal cells; diamorphine is another drug that fits into this category
- to allow the drug to dissolve in water more easily – e.g. fosphenytoin
- to target drugs to a particular area – for example, omeprazole again, where the active drug is formed only in the highly acidic conditions of the cells in the stomach lining.

From this it can be seen that a knowledge of the biochemical processes that occur in the body is essential when designing drugs that are to be converted to an active metabolite in the body.

Buffer solutions

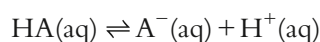
Buffers are important both in the formulation of certain drugs and also most of the reactions that occur in the body do so in aqueous environments where the pH is carefully controlled.

A buffer solution is one that resists changes in pH when small amounts of acid or alkali are added.

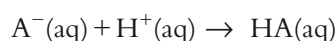
The graph in Figure D.23 shows the result of adding 10 cm³ of 0.100 mol dm⁻³ hydrochloric acid in stages to 100 cm³ of water (blue line) and to 100 cm³ of a buffer solution (orange line).

A buffer solution consists of two components – an acid and a base. The base reacts with any acid added and the acid reacts with any base added. There must be reasonably large amounts of each present for the solution to function as a buffer.

Consider a general buffer containing acid, HA and base A⁻. The equilibrium that exists in this solution is:

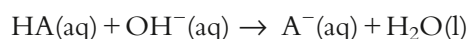


If some hydrochloric acid is added to this solution, the extra H⁺ added reacts with the A⁻ (base) in the solution:



The H⁺ added is ‘mopped up’ by reaction with the base and therefore the pH changes very little.

If some sodium hydroxide is added to the solution, the extra OH⁻ added reacts with the HA (acid) in the solution:



The OH⁻ added is ‘mopped up’ by reaction with the acid and, once again, the pH changes very little.

Buffers can only be made from a weak acid and its conjugate base or a weak base and its conjugate acid – the acid and base present in the buffer must always be a conjugate pair. Buffers cannot be made from a strong acid and its conjugate base or a strong base and its conjugate acid. The strong acid, for example, will be completely dissociated in solution and its conjugate base will have very little tendency to pick up protons when more acid is added.

Buffers contain weak acids – a weak acid is one that dissociates partially in aqueous solution. pK_a provides a measure of how much it dissociates – the smaller the value of pK_a, the more the acid dissociates and the stronger it is. pK_a is different for different acids and also varies with temperature. pK_a is discussed in more detail in the Higher Level section of Topic 8 (Subtopic 8.7).

How to calculate the pH of a buffer solution

For a buffer solution made up of a mixture of HA(acid) and A⁻(base), the pH of the buffer can be worked out by using the **Henderson–Hasselbalch equation**:

$$\text{pH} = \text{pK}_a + \log_{10} \left(\frac{[\text{A}^{-}]}{[\text{HA}]}\right)$$

Higher Level students will have already met the idea of a buffer solution in Topic 8.

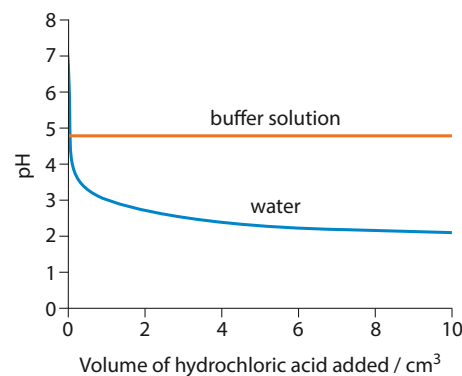


Figure D.23 The orange line shows the effect of adding hydrochloric acid to 100 cm³ of buffer solution formed by mixing 50 cm³ of 1.00 mol dm⁻³ ethanoic acid and 50 cm³ of 0.100 mol dm⁻³ sodium ethanoate.

pK_a works a little like pH (the ‘p’ has the same meaning) – the higher [H⁺(aq)], the lower the pH.

Another way of writing this is:

$$\text{pH} = \text{p}K_{\text{a}} + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

Worked example

D.3 Calculate the pH of a buffer solution containing $0.0550 \text{ mol dm}^{-3}$ CH_3COOH ($\text{p}K_{\text{a}} = 4.76$) and $0.0450 \text{ mol dm}^{-3}$ CH_3COO^- .

$$[\text{base}] = 0.0450 \text{ mol dm}^{-3}; [\text{acid}] = 0.0550 \text{ mol dm}^{-3}$$

$$\text{pH} = \text{p}K_{\text{a}} + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$\text{pH} = 4.76 + \log_{10} \left(\frac{0.0450}{0.0550} \right)$$

$$= 4.76 + \log_{10} 0.818$$

$$= 4.76 - 0.0872$$

$$= 4.67$$

Calculating the pH of a buffer solution when volumes are given

Worked examples

D.4 A buffer solution is formed when 30.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ potassium dihydrogen phosphate (KH_2PO_4) is added to 40.0 cm^3 of $0.110 \text{ mol dm}^{-3}$ disodium hydrogen phosphate (Na_2HPO_4). $\text{p}K_{\text{a}}$ for H_2PO_4^- is 7.21. Calculate the pH of the mixture.

A buffer solution is made up in aqueous solution and so the salts will be split apart into ions; therefore the solution contains the dihydrogen phosphate ion (H_2PO_4^-) and the hydrogen phosphate (HPO_4^{2-}) ion. H_2PO_4^- has an extra proton and acts as an acid, whereas HPO_4^{2-} , which has one less proton, can act as a base. The potassium ions and sodium ions are not important for the working of the buffer – they are there because you cannot have a solution containing just negative ions – it has to be neutral overall.

The first step is to work out the concentrations of the acid and base in the buffer solution.

The total volume of the solution is 70.0 cm^3 . Because the same number of moles of potassium dihydrogen phosphate are now present in 70.0 cm^3 instead of 30.0 cm^3 , the concentration of the potassium dihydrogen phosphate has decreased by a factor of $\frac{30}{70}$.

The concentration of potassium dihydrogen phosphate in this solution will be:

$$\left(\frac{30.0}{70.0} \right) \times 0.100 = 0.0429 \text{ mol dm}^{-3}$$

The concentration of disodium hydrogen phosphate in this solution will be:

$$\left(\frac{40.0}{70.0} \right) \times 0.110 = 0.0629 \text{ mol dm}^{-3}$$

Exam tip

The species with more H atoms will be the acid (HA); the species with fewer H atoms or more negative charge or less positive charge will be the base (A^-).

More figures were carried through on the calculator to give this answer.

So $[\text{base}] = 0.0629 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.0429 \text{ mol dm}^{-3}$

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$\text{pH} = 7.21 + \log_{10} \left(\frac{0.0629}{0.0429} \right)$$

$$= 7.21 + \log_{10} 1.47$$

$$= 7.21 + 0.166$$

$$= 7.38$$

Exam tip

You can check whether your answer for working out the pH of a buffer solution is reasonable – if the solution contains a higher concentration of acid than base, the pH of the solution will be lower than the $\text{p}K_a$ of the acid; if there is a higher concentration of base than acid, the pH will be higher than the $\text{p}K_a$.

The concentration of each species in the buffer solution can also be worked out using a moles calculation.

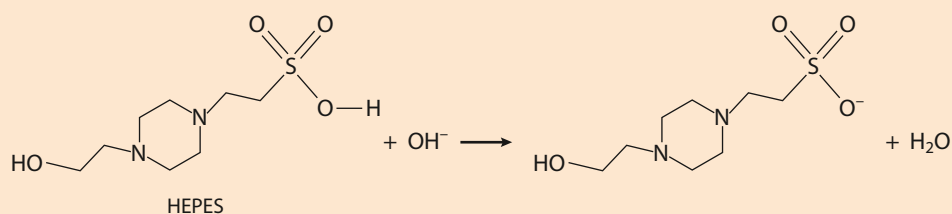
The number of moles of potassium dihydrogen phosphate in 30.0 cm^3 :

$$\left(\frac{30.0}{1000} \right) \times 0.100 = 0.00300 \text{ mol}$$

So the concentration of potassium dihydrogen phosphate in the buffer solution is:

$$\left(\frac{0.00300}{70} \right) \times 1000 = 0.0429 \text{ mol dm}^{-3}$$

D.5 HEPES is used in some biological buffers. A buffer solution can be made by dissolving sodium hydroxide in a HEPES solution.



Calculate the pH of the buffer solution formed when 20.0 g of sodium hydroxide is added to 1.00 dm^3 of a 1.00 mol dm^{-3} solution of HEPES ($\text{p}K_a = 7.5$). Assume that there is no change in volume when the sodium hydroxide is added.

HEPES has an 'extra' proton and is therefore an acid. Reaction with sodium hydroxide converts some of it into a base.

M_r for sodium hydroxide is 40.00

$$\text{So the number of moles of sodium hydroxide} = \frac{20.0}{40.00} = 0.500 \text{ mol.}$$

From the equation, there is a 1:1 reaction with sodium hydroxide and therefore 0.500 mol HEPES reacts with 0.500 mol NaOH to form 0.500 mol of the anion.

In 1.00 dm^3 of a 1.00 mol dm^{-3} solution of HEPES there is 1.00 mol of HEPES. So if 0.500 mol react there will be 0.500 mol remaining. Therefore the concentration of HEPES and the anion in the buffer solution are both equal at $0.500 \text{ mol dm}^{-3}$.

$$[\text{base}] = 0.500 \text{ mol dm}^{-3}; [\text{acid}] = 0.500 \text{ mol dm}^{-3}$$

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$\text{pH} = 7.5 + \log_{10} \left(\frac{0.500}{0.500} \right)$$

$$= 7.5 + \log_{10} 1$$

$$= 7.5 + 0$$

$$= 7.5$$

Determining the composition of a buffer solution given its pH

Worked examples

D.6 A student wants to make up a buffer solution of pH 7.7 using $0.100 \text{ mol dm}^{-3}$ solutions of HEPES ($\text{p}K_a = 7.5$) and its sodium salt. Calculate how much of each solution must be used to make 500 cm^3 of a buffer of pH 7.7.

We need to calculate the ratio of the acid and base in the buffer solution – this can be worked out using the Henderson–Hasselbalch equation.

$$\text{pH} = \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$7.7 = 7.5 + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

$$\log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right) = 0.2$$

$$\frac{\text{base}}{\text{acid}} = 10^{0.2} = 1.58$$

10^x is the inverse function of \log_{10} – use the key combinations ‘shift log’ or ‘2nd log’ on your calculator.

Therefore the ratio $[\text{base}] : [\text{salt}]$ is $1.58 : 1$

Because the concentrations of the solutions are the same, the amount of each solution required to make 500 cm^3 of buffer can be worked out as:

$$\text{volume of base} = \left(\frac{1.58}{2.58} \right) \times 500 = 306 \text{ cm}^3$$

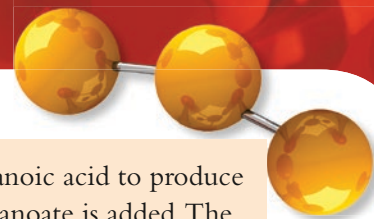
2.58 is $1.58 + 1$ from the ratio

$$\text{volume of acid} = \left(\frac{1.00}{2.58} \right) \times 500 = 194 \text{ cm}^3$$

This could also be worked out using $500 - 306$.

Therefore the volume of the HEPES solution required is 194 cm^3 and that of the solution of its sodium salt is 306 cm^3 .

If all figures are carried through on the calculator the answers 193 cm^3 and 307 cm^3 are obtained.



D.7 What mass of solid sodium ethanoate must be added to 100.0 cm^3 of 0.200 mol dm^{-3} ethanoic acid to produce a buffer solution of pH 4.00? Assume there is no change in volume when the sodium ethanoate is added. The pK_a for ethanoic acid is 4.76.

$$\text{pH} = pK_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}\right)$$

The base is the ethanoate ion (CH_3COO^-) and the acid is ethanoic acid (CH_3COOH)

$$\text{So, } 4.00 = 4.76 + \log_{10} \left(\frac{[\text{CH}_3\text{COO}^-]}{[\text{CH}_3\text{COOH}]}\right)$$

$$\log_{10} \left(\frac{[\text{CH}_3\text{COO}^-]}{[\text{CH}_3\text{COOH}]}\right) = 4.00 - 4.76 = -0.76$$

$$\left(\frac{[\text{CH}_3\text{COO}^-]}{[\text{CH}_3\text{COOH}]}\right) = 10^{-0.76} = 0.174$$

The concentration of ethanoic acid is 0.200 mol dm^{-3} . Substituting this into the above equation we get:

$$[\text{CH}_3\text{COO}^-] = 0.174 \times 0.200 = 0.0348\text{ mol dm}^{-3}$$

So, the concentration of the sodium ethanoate in the solution must be $0.0348\text{ mol dm}^{-3}$, but because we are making up 100 cm^3 of the buffer solution, the number of moles of sodium ethanoate that must be used is given by:

$$\text{no. moles} = \text{concentration} \times \text{volume (in dm}^3\text{)}$$

$$= 0.0348 \times \left(\frac{100.0}{1000} \right)$$

$$= 0.00348\text{ mol}$$

The molar mass of sodium ethanoate is 82.04 g . Therefore:

$$\text{mass} = \text{no. moles} \times \text{molar mass}$$

$$= 0.00348 \times 82.04$$

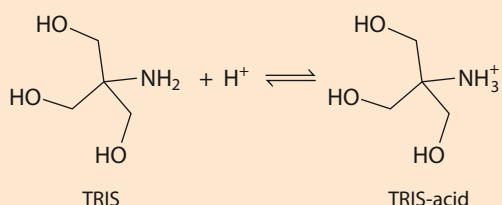
$$= 0.285\text{ g}$$

So 0.285 g of sodium ethanoate must be dissolved in the ethanoic acid to produce a buffer solution of $\text{pH} = 4.00$.

Calculating the change in pH of a buffer solution when acid or alkali is added

Worked example

D.8 TRIS is used as a buffer in biochemistry. A buffer solution is prepared by adding hydrochloric acid to TRIS to form a mixture of TRIS and its protonated form (TRIS-acid). The equilibrium that exists in the buffer solution is:



- a** Calculate the pH of a buffer solution containing $0.750 \text{ mol dm}^{-3}$ TRIS-acid ($\text{p}K_{\text{a}} = 8.30$) and $0.750 \text{ mol dm}^{-3}$ TRIS.
- b** What is the pH of the solution formed when 10.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ hydrochloric acid is added to 50.0 cm^3 of the buffer solution in part **a**?

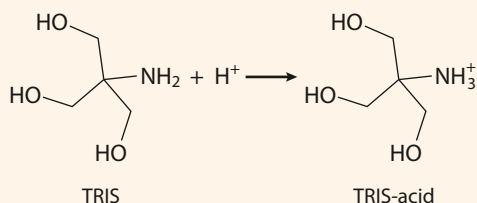
a $[\text{base}] = 0.750 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.750 \text{ mol dm}^{-3}$

$$\begin{aligned} \text{pH} &= \text{p}K_{\text{a}} + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]}} \right) \\ &= 8.30 + \log_{10} \left(\frac{0.750}{0.750} \right) \\ &= 8.30 \end{aligned}$$

Exam tip

When $[\text{base}] = [\text{acid}]$, the pH of the buffer is equal to the $\text{p}K_{\text{a}}$ of the acid.

- b** When some acid is added to the buffer solution, the following reaction occurs:



This means that the concentration of TRIS decreases and that the concentration of TRIS-acid increases. To work out by how much they change we need to work out the initial number of moles of TRIS and TRIS-acid and how many moles of acid were added.

The number of moles of TRIS in 50.0 cm^3 of $0.750 \text{ mol dm}^{-3}$ solution is given by:

$$\begin{aligned} \text{no. moles} &= \text{concentration} \times \text{volume in dm}^3 \\ &= 0.750 \times \left(\frac{50.0}{1000} \right) \\ &= 0.0375 \text{ mol} \end{aligned}$$

This is the same as the number of moles of TRIS-acid.



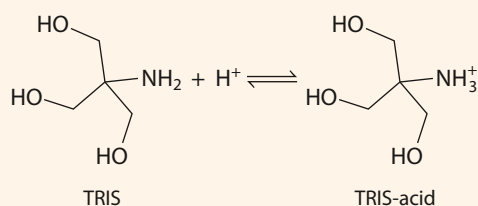
The number of moles of hydrochloric acid in 10.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ solution is given by:

$$\text{no. moles} = 0.100 \times \left(\frac{10.0}{1000} \right)$$

$$= 1.00 \times 10^{-3} \text{ mol.}$$

We will assume that the H^+ from the hydrochloric acid reacts with the equivalent amount of TRIS and that there is no further change in the number of moles as equilibrium is established.

So the number of moles of TRIS thus decreases by $1.00 \times 10^{-3} \text{ mol}$ and the number of moles of TRIS-acid increases by $1.00 \times 10^{-3} \text{ mol}$.



Initial amounts / mol:	0.0375	0.0375 mol
After HCl added / mol:	$0.0375 - 1.00 \times 10^{-3}$ $= 0.0365$	$0.0375 + 1.00 \times 10^{-3}$ $= 0.0385$

The concentration of each species can be worked out by dividing the number of moles by the total volume in dm^3 , which is $50.0 + 10.0 = 60.0 \text{ cm}^3$ or 0.0600 dm^3

Concentration / mol dm^{-3} :	$\frac{0.0365}{0.0600}$ $= 0.608$	$\frac{0.0385}{0.0600}$ $= 0.642$
--	--------------------------------------	--------------------------------------

So $[\text{base}] = 0.608 \text{ mol dm}^{-3}$; $[\text{acid}] = 0.642 \text{ mol dm}^{-3}$

$$\begin{aligned} \text{pH} &= \text{p}K_a + \log_{10} \left(\frac{[\text{base}]}{[\text{acid}]} \right) \\ &= 8.30 + \log_{10} \left(\frac{0.608}{0.642} \right) \\ &= 8.28 \end{aligned}$$

So, upon addition of 10.0 cm^3 of the hydrochloric acid, the pH of the buffer solution falls by 0.02 to 8.28.

Nature of science

Science is a highly collaborative enterprise and the development of a drug involves scientists with many different specialisms. It is also important that individual scientists have knowledge outside their particular specialism – for instance, a knowledge of both organic chemistry and biochemistry was essential in the development of drugs such as omeprazole and esomeprazole.

Science is an ever-changing body of knowledge and collection and analysis of data is an important part of developing new theories. It used to be thought that stress and increased stomach acid were major causes of stomach ulcers until Warren and Marshall discovered that most stomach ulcers are actually caused by a bacterial infection. They were awarded a Nobel Prize in 2005.

? Test yourself

- 2 Calculate the pH values of the following buffer solutions:
 - a A solution containing $0.0200 \text{ mol dm}^{-3}$ butanoic acid ($\text{p}K_{\text{a}} = 4.82$) and $0.0200 \text{ mol dm}^{-3}$ sodium butanoate.
 - b A solution containing $0.0500 \text{ mol dm}^{-3}$ propanoic acid ($\text{p}K_{\text{a}} = 4.87$) and $0.0200 \text{ mol dm}^{-3}$ sodium propanoate.
 - c A solution containing $0.300 \text{ mol dm}^{-3}$ ethanoic acid ($\text{p}K_{\text{a}} = 4.76$) and $0.500 \text{ mol dm}^{-3}$ sodium ethanoate.
 - d A solution made up by mixing together 25.0 cm^3 of $0.200 \text{ mol dm}^{-3}$ ethanoic acid ($\text{p}K_{\text{a}} = 4.76$) and 50 cm^3 of $0.100 \text{ mol dm}^{-3}$ sodium ethanoate.
 - e A solution obtained when 10.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ sodium hydroxide is added to 20.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ ethanoic acid ($\text{p}K_{\text{a}} = 4.76$).
 - f A solution obtained when 20.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ ammonia solution is added to 40.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ ammonium chloride solution ($\text{p}K_{\text{a}} = 9.25$).
- 3
 - a A buffer solution contains ethanoic acid ($\text{p}K_{\text{a}} = 4.76$) at a concentration of 1.00 mol dm^{-3} and sodium ethanoate. If the pH of the buffer solution is 4.20, what is the concentration of the sodium ethanoate?
 - b 20.0 cm^3 of $0.0100 \text{ mol dm}^{-3}$ hydrochloric acid is added to 50 cm^3 of the buffer solution in part a. Calculate the new pH of the buffer solution.
- 4 What mass of solid sodium propanoate must be added to 50.0 cm^3 of $0.100 \text{ mol dm}^{-3}$ propanoic acid ($\text{p}K_{\text{a}} = 4.87$) to produce a buffer solution of pH 5.00? Assume there is no change in volume when the sodium propanoate is added.

Learning objectives

- Describe the main differences between viruses and bacteria
- Understand the various modes of action of antivirals
- Understand how oseltamivir (Tamiflu[®]) and zanamivir (Relenza[™]) work
- Compare the structures of oseltamivir and zanamivir
- Discuss why it is difficult to solve the global AIDS problem

D5 Antiviral medications

Viruses

Viruses are **parasites** – they invade host cells and use the materials and processes within those cells to produce new viruses – they cannot replicate outside host cells.

Viruses differ greatly in shape and size from one type to the next but, in general, they have a core consisting of their genetic information (carried in the form of either DNA or RNA) which is surrounded by a protein coat known as a **capsid**. This capsid consists of identical protein subunits, called capsomeres, and its role is to protect the genetic information in the core. The capsid and genetic material together are called a **nucleocapsid**. Some viruses, such as the human immunodeficiency virus (HIV), also have a lipid envelope that surrounds the nucleocapsid (Figure D.24).

Viruses are not considered to be living cells – they do not feed, excrete or grow and they consist only of what is necessary to invade the host cell and then take over that cell to produce copies of themselves. On the other hand, bacteria are living cells and are far more complex in structure and function than viruses – they are also able to reproduce outside host cells by cell division.

Some differences between viruses and bacteria are summarised in Table D.1.

Bacteria	Viruses
living	not living
larger than viruses	smaller than bacteria
have a cellular structure – cell wall, cytoplasm, nucleus etc.	do not have a cellular structure
contain DNA and RNA	contain either DNA or RNA
can reproduce independently	need a host cell to reproduce

Table D.1 Differences between viruses and bacteria.

The lack of cell structure makes it much more difficult to design drugs to target viruses than bacteria.

When designing drugs, there are far fewer enzymes/receptor proteins to target in viruses – viruses mostly use enzymes belonging to the host cell. Bacteria can be targeted, for instance by inhibiting the enzymes that make cell walls, but viruses don't have cell walls.

To gain entry into host cells, viruses must first attach to the surface of a host cell. The genetic material of the virus is released into the cytoplasm and is then incorporated into the host cell's DNA (if the virus contains RNA this must first be converted into DNA before it is inserted). The cell then starts producing viral proteins and viral DNA or RNA which get assembled into functional new viruses that leave the cell to go on to infect other cells.

Treatment and prevention of viral diseases

Viruses cause a number of illnesses and diseases ranging from mild infections, such as the common cold, to potentially fatal diseases, such as acquired immunodeficiency syndrome (AIDS). It can be difficult to find effective methods of preventing and treating viral infections for a number of reasons:

- Once inside a host cell, a virus can multiply very quickly and can have already spread throughout the body by the time that symptoms have appeared.
- Viruses can mutate their DNA or RNA resulting in a slight change in viral structure – this can make them resistant to drugs and can prevent vaccinations from being effective; this is particularly true of viruses such as HIV.
- Viruses use the host cell's own processes and materials to produce new viruses, so it can be difficult to design drugs that target only the virus and do not affect the host cell.

However, despite these difficulties, several **vaccines** and **antiviral drugs** have been developed and used to prevent and treat viral infections successfully.



Figure D.24 The HIV virus.

Vaccines stimulate the body's natural defences (the immune system) to produce antibodies against a virus – so if infection does occur, the immune system is prepared and can stop the infection before it takes hold. Vaccines have been used successfully against a number of viruses including measles, mumps and polio.

Antiviral drugs work in a number of ways.

- Some **alter the genetic material** within cells – once inside a cell, the drug is converted into an active metabolite that becomes incorporated into the growing DNA strand (needed for viral replication) halting its synthesis. An example of a drug that acts in this way is aciclovir (acyclovir), which is used to treat cold sores; it stops viral DNA replication and so stops the virus from multiplying.
- Some **inhibit the activity of enzymes** within the host cell that are necessary for the formation of new viruses. An example is indinavir, which is used in AIDS treatment; it inhibits the HIV enzyme protease, which is essential to the assembly of functional new HIV viruses.
- Some stop the viruses from infecting host cells by **preventing them from binding to the host cell surface** and gaining access into the cell. Some drugs used to treat AIDS work in this way.
- Some **prevent the virus from leaving the host cell** so that it cannot infect other cells – see oseltamivir and zanamivir below.

Influenza antivirals

Influenza ('flu') is a viral disease that most people are familiar with – its symptoms include fever, headache, aching joints and fatigue. In severe cases it can cause death and there have been many cases of flu pandemics such as the outbreak of so called 'Spanish flu' in 1918, which killed millions of people, and 'swine flu' in 2009, which killed thousands.

The influenza virus has RNA as its genetic material rather than DNA. The capsid contains RNA and RNA polymerase (an enzyme that produces a type of viral RNA which is used by the cell to make viral proteins) and surrounding this it has a lipid envelope (derived from the host cell membrane) which has two very important proteins in it – **neuraminidase** (NA) and **hemagglutinin** (HA). Both proteins stick out of the membrane surrounding the virus. Hemagglutinin allows viruses to bind to receptors on the surface of cells to be infected – it binds to sialic acid residues on glycoproteins (proteins joined to sugar groups). Neuraminidase is an enzyme that hydrolyses the bond between a sialic acid residue and the rest of the glycoprotein – it essentially adds water across the bond to break it. Neuraminidase is important for helping the viruses to infect cells in that it breaks down the mucous surrounding cells in the upper respiratory tract, allowing the virus access to the target cell. Neuraminidase also plays an essential role at the end of the virus lifecycle – the virus enters the cell, replicates its RNA and essential proteins, but then the new viruses have to get out of the cell if they are to infect other cells. They do this by budding (Figure D.25). The virus will, however, remain bound to the surface of the cell (because the HA protein binds to receptors on the surface) until the NA breaks the link between the sialic acid and the rest of the membrane glycoprotein, allowing the virus to break free from the cell.

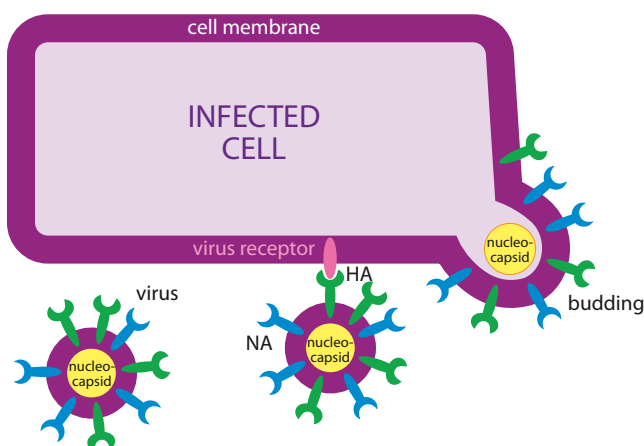


Figure D.25 Viruses must escape from the host cell.

Influenza antiviral drugs such as oseltamivir (Tamiflu) and zanamivir (Relenza) are **neuraminidase inhibitors**. They bind to the active site of the neuraminidase enzyme, which prevents it from catalysing the hydrolysis of the sialic acid residue from the glycoproteins in the cell membrane, so that viruses remain anchored to the cell and cannot infect other cells.

Oseltamivir/zanamivir may be taken when flu symptoms develop but they are also taken as a prophylactic (a preventative measure) for people who have been exposed to the virus or are in high-risk groups (such as very young children).

Oseltamivir is administered orally as oseltamivir phosphate but zanamivir has very poor bioavailability when given orally and so is formulated as a dry powder for oral inhalation. Zanamivir is highly polar (see below) and cannot pass across cell membranes.

Comparison of the structures of oseltamivir and zanamivir

The structures of oseltamivir and zanamivir are compared in Figure D.26. There are certain similarities between the structures – e.g. the six-membered ring at the centre (in oseltamivir this is a cyclohexene ring but in zanamivir it could be called a dihydropyran ring). They also both contain a secondary carboxamide (amide) group. The ester and carboxyl (carboxylic acid) groups are highlighted in the same colour in Figure D.26 because the ester group in oseltamivir is hydrolysed in the body, converting it to a carboxyl group (see ‘active metabolites’ on page 26). Similarities between the general structures would be expected because they both bind to the same active site in the viral neuraminidase enzyme. There are also lots of differences between the structures highlighted in Figure D.26. The group highlighted grey in the zanamivir structure has been broken down into its constituent parts because they are more familiar from earlier work in the Coursebook but it actually constitutes a **guanidine group** (much as an ester is not really a ketone group and an ether group).

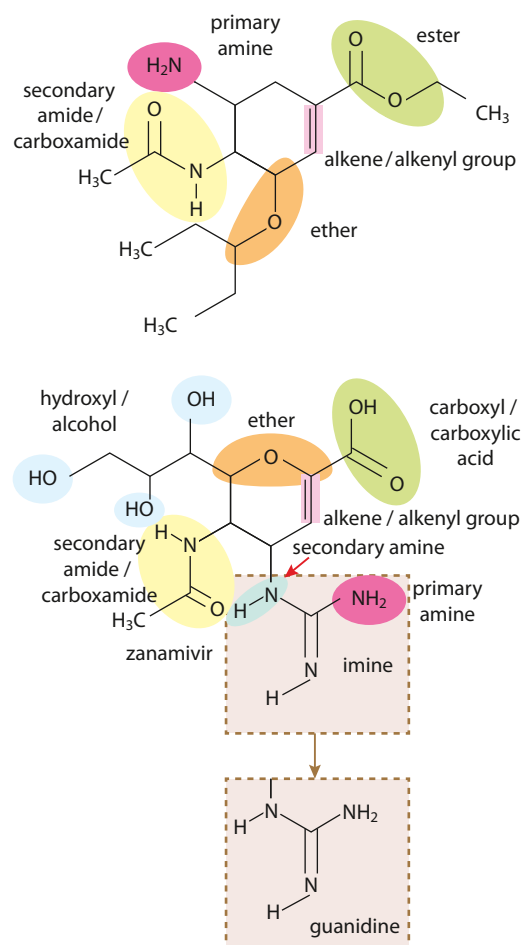


Figure D.26 Comparing the structures of oseltamivir and zanamivir.

AIDS

AIDS was first recognised in 1981 and was found to be caused by the HIV virus a few years later. There are now believed to be more than 34 million people infected with HIV worldwide, and approximately 2 million deaths occur each year from AIDS.

HIV is a retrovirus and its genetic information is carried in the form of RNA, not DNA. HIV infection is lethal if left untreated because the HIV virus invades cells that form part of the immune system – these cells are white blood cells known as **T-cells** and they play a vital role in the body's natural defence against infection. The HIV virus is able to infect these T-cells because they have **specific receptor proteins** on their surface to which the **virus attaches** to gain entry into the cell. Once inside, the viral enzyme **reverse transcriptase** converts the viral RNA into DNA so that it can be integrated into a T-cell's DNA. The viral genes contained in the DNA are used to produce viral proteins and viral RNA within the cell, and these get assembled into new HIV viruses. The T-cell stops carrying out its role as an immune cell and instead becomes a factory for HIV viruses. When the newly formed viruses leave the T-cell, some of the T-cell membrane forms an envelope around the HIV virus.

Death of T-cells can occur when viruses exit a cell and this results in a decrease in the number of T-cells in the body and a weakened immune system. This is why people with AIDS are susceptible to potentially fatal infections and also some types of cancer – their immune systems are not strong enough to fight against them.

Developing a method of eradicating HIV is difficult because the virus is able to mutate rapidly, and also because the virus uses many of the host cell's processes and materials to replicate – so it is difficult to target the virus without affecting the host cell too. However, there are some differences that can be targeted. The HIV virus uses certain viral enzymes in the replication process that are different to those found in the host cell. One of these enzymes is called reverse transcriptase, and the first antiretroviral drug that came onto the market to treat AIDS, azidothymidine (AZT), acts by targeting this enzyme. AZT inhibits reverse transcriptase and gets incorporated into the DNA strand that is being synthesised by the enzyme – this results in termination of DNA synthesis, and so the virus cannot replicate. Other viral enzymes that are inhibited by drugs include the one that integrates the DNA into the host cell's DNA (called integrase) and the one that assembles the viral proteins to produce new viruses (called protease). Drugs are also available that stop the virus from binding to T-cell receptor proteins and gaining entry into the host cell. However, all these drugs only delay the progression of AIDS, they do not destroy the virus – nevertheless they have saved the lives of millions of people since their introduction.

Much research is being undertaken to find an effective vaccine that can be used against HIV to try to stop the spread of the virus. The ability of the virus to mutate and change its structure has made it difficult to find a suitable vaccine that can prime the host's natural immunity against such variations in structure. Promising results are being reported by researchers, both in animal studies and human trials, where a reduction in infection rates has been shown.



One problem with antiretroviral therapy is that it used to be expensive, so AIDS sufferers in poor countries (where the majority of AIDS cases are found) did not generally have access to these life-saving drugs. However, the prices of the most commonly used antiretroviral treatments have decreased significantly over the last few years and, together with a global commitment to make these treatments universally available, more and more patients in poorer countries are now receiving treatment. However, more work still needs to be done to ensure that prevention measures (such as education and condoms) and antiretroviral treatments are available to all.



Nature of science

Science can be used for good and for bad and there are many ethical issues that face scientists working in certain fields. As well as medical programs that have developed vaccines for smallpox and cured many other diseases, there are many rumours about biological weapons employing viruses or bacteria in top-secret programs in various countries around the world. There have also been cases of terrorist attacks using anthrax bacteria.

Science is a highly collaborative activity. Advances in our understanding of how viruses infect cells and the development of drugs to treat viral infections have involved many different people with a variety of specialisms working together.

D6 Environmental impact of some medications

Nuclear waste

There are many applications of radioactive isotopes in medicine – they are used for both treatment and diagnosis. Radioactive isotopes undergo radioactive decay by the emission of alpha particles, beta particles or gamma rays (there are other forms of radioactive decay). These are called **ionising radiation** because they cause the formation of ions (by ejection of electrons) when they interact with matter. Ionising radiation can damage cells and the main effect comes from damage caused to DNA.

Numerous radioisotopes are used or produced in various medical processes and these have half-lives that vary enormously – ^{131}I , used to treat thyroid cancer, has a half-life of just 8 days; ^{60}Co , used to treat other forms of cancer, has a half-life of 5.3 years.

Radioactive waste is a byproduct of the use of radioisotopes in medicine and, of course, it must be disposed of in some way. The classification of radioactive waste is important in determining how it can be disposed of and the safety measures that must be used in its transport and handling.

Radioactive waste can be divided into different categories – low-level and high-level. The category ‘intermediate-level radioactive waste’ is also sometimes used. The criteria used for the classification are quite complex and low-level waste can be divided into sub-categories (A, B, C and >C) depending on the level of activity.

- Low-level waste has a low activity (not many radioactive nuclei decay each second to produce ionising radiation) and usually contains isotopes with short half-lives (ionising radiation is given off for a shorter period of time).
- High-level waste has a high activity (many radioactive nuclei decay each second to produce ionising radiation) and usually contains isotopes with longer half-lives (ionising radiation is given off for a long time).

One of the ways of reducing the environmental impact of nuclear medicine is to choose the radioactive sources very carefully – they should

Learning objectives

- Distinguish between high-level and low-level nuclear waste
- Discuss the environmental effects of medical nuclear waste
- Explain the dangers of antibiotic waste
- Understand the basics of green chemistry
- Understand how green chemistry was involved in the production of a precursor of oseltamivir
- Discuss the environmental issues associated with the use of solvents in the pharmaceutical industry

Half-life is the time it takes for the number of radioactive nuclei present in a sample at any given time to fall to half its value.

have the minimum possible activity and the shortest possible half-lives, but still be suitable for the job.

Radioactive materials have the potential to be a serious hazard both to people and to the environment – therefore the disposal of medical nuclear waste must be controlled carefully. Governments set strict limits on the release of radioactivity into the environment and monitoring is essential to ensure that these regulations are adhered to. The main approaches to the disposal of nuclear waste are ‘dilute and disperse’, ‘delay and decay’ or ‘confine and contain’. ‘Confine and contain’ is always used for nuclear waste that has a high level of activity.

Health effects that can result from long-term exposure to increased low-level radioactivity, for example, from the disposal of medical nuclear waste, include cancer and mutations in DNA.

Although the waste from medical processes is generally classified as low-level, some radioisotopes that are used in medicine are produced in nuclear reactors, which generate high-level waste.

Low-level waste

This includes items that have been contaminated with radioactive material or have been exposed to radioactivity. Examples are gloves and other protective clothing, tools, syringes and excreta from patients treated with radioisotopes.

Low-level waste may be stored on site until it has decayed to such an extent that it can be disposed of as ordinary waste (e.g. in landfill sites or released into the sewage system) or shipped to a central site for more specialised disposal.

Some low-level waste is incinerated, which reduces its volume considerably and distributes the radioisotopes over a wide area – ‘dilute and disperse’. The ash from incineration is assessed for activity and disposed of appropriately. Low-level waste with higher activity is often just buried underground (‘near-surface’ disposal) – for example, in individual concrete canisters or in concrete-lined vaults. Low-level waste may need to be contained underground for up to 500 years depending on its activity and half-life.

High-level waste

This includes spent fuel rods and other materials from nuclear reactors. High-level waste will remain hazardous to humans and other living things for thousands of years.

High-level liquid waste can be converted to glass (vitrification) to make storage easier. High-level waste is first kept in storage pools (cooling ponds) under water, usually for a minimum of nine months, but sometimes spent fuel rods are stored in this way for decades. After sufficient cooling, fuel rods may be transferred to dry storage casks – these have very thick walls and are made of steel and concrete. The dry casks are then stored in concrete bunkers.

Permanent storage of high-level radioactive waste is a major problem and various solutions have been suggested – such as burying the waste deep underground in stable geological areas. Over thousands of years, however, it is difficult to predict what processes could occur to cause release of the radioactive material. Many people argue that there is no suitable solution for the disposal of high-level waste.



The release of antibiotics into the environment

The presence of antibiotics and other pharmaceuticals in waste water is becoming an increasing problem. Antibiotics can enter the water supply by several routes. These include:

- incorrect disposal of unwanted medicines – for example, by flushing old medicines down the toilet
- agriculture – drugs given to animals will be present in animal waste (urine and feces) and can find their way into groundwater, rivers and lakes.

Treating water to produce drinking water does remove some of these chemicals but there is still a variety of pharmaceuticals present in the water we drink. Although these pharmaceuticals are found in drinking water in only very small amounts (typically nanograms per dm^3) there are concerns that long-term exposure could result in damage to human health.

The release of antibiotics into the environment is regarded as a particular problem because not only can they cause damage to aquatic organisms, but they can also result in increased resistance of bacteria to antibiotics. Antibiotics (antibacterials) are used to treat a variety of conditions but if bacteria develop resistance to antibiotics such as penicillin, these diseases can become much more difficult to cure. Antibiotic resistance is discussed more fully on page 17.

Green chemistry

Green chemistry (or sustainable chemistry) is an approach to chemical research and chemical industrial processes that seeks to minimise the production of hazardous substances and their release into the environment.

There are 12 principles of green chemistry, an idea developed by Paul Anastas and John C. Warner:

- 1 *Prevention* – it is better to prevent waste than to treat or clean up waste after it has been created.
- 2 *Atom economy* – synthetic methods should be designed to maximise the incorporation of all materials used in the process into the final product.
- 3 *Less hazardous chemical syntheses* – wherever practicable, synthetic methods should be designed to use and generate substances that possess little or no toxicity to human health and the environment.
- 4 *Designing safer chemicals* – chemical products should be designed to affect their desired function, while minimising their toxicity.
- 5 *Safer solvents and auxiliaries* – the use of auxiliary substances (solvents, separation agents etc.) should be made unnecessary wherever possible and innocuous when used.
- 6 *Design for energy efficiency* – the energy requirements of chemical processes should be recognised for their environmental and economic impacts and should be minimised. If possible, synthetic methods should be conducted at ambient temperature and pressure.
- 7 *Use of renewable feedstocks* – a raw material or feedstock should be renewable, rather than depleting, whenever technically and economically practicable.

- 8 *Reduce derivatives* – unnecessary derivatisation (use of blocking groups, protection/ deprotection, temporary modification of physical/ chemical processes) should be minimised or avoided if possible because such steps require additional reagents and can generate waste.
- 9 *Catalysis* – catalytic reagents (as selective as possible) are superior to stoichiometric reagents.
- 10 *Design for degradation* – chemical products should be designed so that at the end of their function they break down into innocuous degradation products and do not persist in the environment.
- 11 *Real-time analysis for pollution prevention* – analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances.
- 12 *Inherently safer chemistry for accident prevention* – substances, and the form of a substance used in a chemical process, should be chosen to minimise the potential for chemical accidents including releases, explosions and fires.

(Green Chemistry: Theory and Practice by Paul Anastas & John Warner (1998) Figure 4.1 from p. 30. By permission of Oxford University Press www.oup.com)

As can be seen above, there are many things that have to be considered when making a pharmaceutical. The best synthetic routes to a drug:

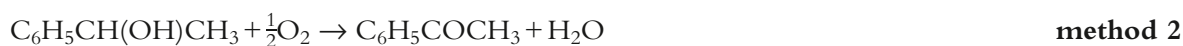
- use readily available and safe materials
- have the minimum number of steps
- convert as much of the starting materials as possible into the required product at each step – good atom economy and good yield
- use as little solvent as possible
- use as little energy as possible.

An important consideration in green chemistry is the concept of atom economy. Atom economy can be used as a measure of how efficient a particular reaction is in terms of converting as much of the starting materials as possible into useful products.

$$\text{atom economy} = \frac{\text{molar mass of desired product}}{\text{total molar mass of all reactants}} \times 100\%$$

We can use the preparation of 1-phenylethanone, which could be investigated as an enzyme inhibitor, as an example to illustrate how the equation is used.

Consider two different ways of making 1-phenylethanone ($\text{C}_6\text{H}_5\text{COCH}_3$), from 1-phenylethanol:



The atom efficiency for each process can be worked out as follows.

Method 1:

$$\begin{aligned} \text{total molar mass of all reactants} &= (3 \times 122.18) + (2 \times 100.00) + (3 \times 98.09) \\ &= 860.81 \text{ g mol}^{-1} \end{aligned}$$

$$\text{molar mass of desired product} = 3 \times 120.16 = 360.48 \text{ g mol}^{-1}$$

$$\text{atom economy} = \left(\frac{360.48}{860.81} \right) \times 100 = 41.88\%$$



Method 2:

$$\text{total molar mass of all reactants} = 122.18 + (0.5 \times 32.00) = 138.18 \text{ g mol}^{-1}$$

$$\text{molar mass of desired product} = 120.16 \text{ g mol}^{-1}$$

$$\text{atom economy} = \left(\frac{120.16}{138.18} \right) \times 100 = 86.96\%$$

It can be seen that method 2 has a much higher atom economy and is, therefore, much more efficient. However, many other things must be considered when assessing these reactions in terms of green chemistry principles – the temperature used in each reaction, the solvents used and how much of them, disposal of the solvents, the nature of the catalyst required for the second reaction etc.

Atom economy is not the same as the yield of a reaction. Atom economy is a theoretical quantity based on a chemical equation and allows an evaluation of how much waste is produced. The yield of a reaction is an experimental quantity worked out from how much of the desired product is actually made in a chemical reaction.

In the calculation of atom economy above, it has been assumed that all reactions have 100% yield which will not be the case in practice.

When evaluating how green/environmentally friendly a particular process is both atom economy and yield must be considered as well as several other factors.

Synthesis of oseltamivir

Oseltamivir has been discussed earlier (page 37) as a treatment for influenza. Total synthesis (from petrochemical starting materials) of oseltamivir involves huge amounts of materials and can generate thousands of kilograms of waste per mole of oseltamivir made. Therefore it was essential to develop greener routes to the drug.

The current commercial synthetic route uses a naturally occurring material, shikimic acid, as the starting material – this cuts out several steps in the synthesis and makes it greener. Shikimic acid is a renewable material that can either be extracted from Chinese star anise or obtained from glucose by fermentation using genetically modified bacteria. Even starting from shikimic acid, a further ten steps are required to make oseltamivir, so more work is required to make this synthesis even greener!

Although shikimic acid can be obtained from star anise, this in itself causes a problem – the yield is not very high and the production of shikimic acid is linked to the availability of star anise. The use of GM bacteria is likely to provide a better long-term solution to producing shikimic acid – fermentation uses relatively low temperatures and an aqueous medium, so it is a reasonably green process.

Waste solvents

There are many environmental issues with the use of solvents in the pharmaceutical industry. Solvents are used as the medium in which many reactions occur, in the extraction and purification of compounds and so on. Solvents contribute typically about 80–90% of the mass of substances used in the production of a pharmaceutical and also make a large contribution to the amount of energy used and the cost. Many of the solvents used

From the point of view of the pharmaceutical industry there is also the problem of residual solvents in drug samples, which need to be removed as far as possible before the drugs can be administered to humans – especially if they are toxic.

are organic because of the nature of the compounds being used and made – usually mostly non-polar and therefore more soluble in organic solvents – and many will be toxic to humans and other organisms.

The first aspect to be considered when trying to make a process greener is prevention – can a different solvent be used that is less harmful to the environment? If a greener solvent cannot be found then the amount used should be reduced as far as possible.

The next consideration is the possibility of a solvent being recycled and reused. This avoids the need to dispose of the solvent. The energy considerations of purifying the solvent for reuse must be considered at this stage.

If the solvent cannot be reused, then it must be disposed of as safely as possible. There are many methods of disposal including incineration and injection underground. The nature of the solvents that are to be injected underground is controlled strictly because the practice could introduce potentially toxic chemicals into the environment. Incineration produces carbon dioxide, a greenhouse gas, which can contribute to climate change and can also produce toxic substances such as dioxins.

Nature of science

An understanding of science is essential if the public are to make informed judgements about the advantages and disadvantages of using antibiotics. It is important that scientists explain the issues in a way that is as complete as possible but also objective so that, for instance, farmers can make informed decisions about adding antibiotics to animal feed etc. and patients can understand the importance of disposing of unwanted medicines in an appropriate way.

When considering whether a particular drug should be licensed for medical use the benefits and risks to patients are considered by regulatory authorities. However, there are also environmental factors that should be considered and scientists are involved in developing processes to reduce the environmental impact of drug development and production.

Learning objectives

- Understand what is meant by a chiral auxiliary
- Understand how chiral auxiliaries can be used for stereoselective syntheses
- Understand that paclitaxel (Taxol[®]) is used in chemotherapy
- Understand how paclitaxel is produced
- Understand how a polarimeter can be used to identify enantiomers

D7 Taxol[®] – a chiral auxiliary case study (HL)

Chirality

When a molecule contains a carbon atom bonded to four different groups, it is said to be chiral and two mirror images (known as enantiomers) exist. These enantiomers can behave very differently in the body as a result of their different shapes. For example, one of the enantiomers may be able to bind effectively to an enzyme or receptor protein because its functional groups are in the correct orientation to form bonds with the protein, whereas the other may not be able to bind as strongly because the groups are in the wrong orientation to form bonds (Figure D.27). In some cases, one enantiomer may produce a therapeutic effect by binding to its target, whereas the other may produce a toxic effect by binding elsewhere.

Thalidomide is an example of a chiral drug that was given as a mixture of enantiomers (a racemic mixture) to pregnant women to combat morning sickness. It was later discovered that one of the enantiomers (the S-enantiomer – Figure D.28) was responsible for producing a teratogenic effect and caused limb deformities in the fetus. Pharmaceutical companies generally now either synthesise or separate out the active single enantiomer of a drug and develop this instead of the racemic mixture.

Administering the single enantiomer would not have helped in the case of thalidomide because the enantiomers interconvert when in the body, producing the racemic mixture. Nowadays, if a new drug is going to be marketed as the racemic mixture, testing for toxicity and effectiveness must be carried out on each enantiomer separately and also on the racemic mixture.

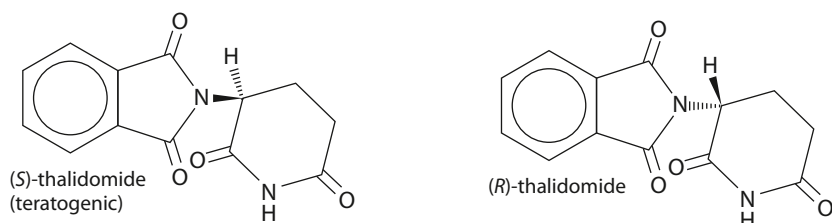


Figure D.28 Enantiomers of thalidomide.

Chiral auxiliaries in asymmetric synthesis

Many drugs have chiral centres and so can exist as two enantiomers, but it is usual for pharmaceutical companies to develop just one enantiomer of a drug for reasons already explained. Synthesis reactions normally produce a mixture of both enantiomers (racemic mixture). Consider Figure D.29, which shows the reaction between a Grignard reagent ($\text{C}_2\text{H}_5\text{MgBr}$) and pentan-2-one. Neither of the starting materials is chiral but the product has a chiral centre (shown green).

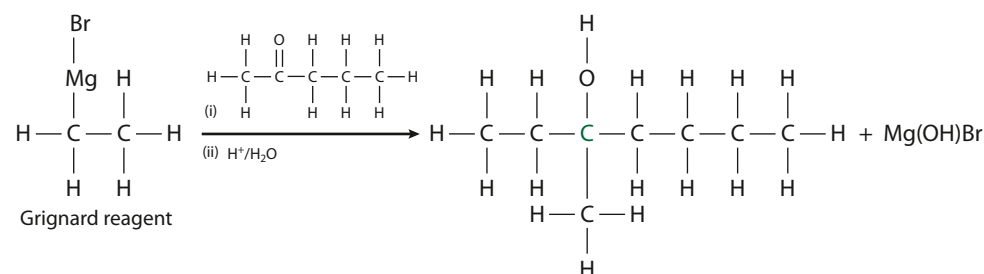


Figure D.29 The reaction of ethylmagnesium bromide with pentan-2-one.

The Grignard reagent reacts as if it is C_2H_5^- and can attack the pentan-2-one (planar about the $\text{C}=\text{O}$ group) from either the top or the bottom. Attack from either side is equally probable and so an equimolar mixture of the two enantiomers is formed – a racemic mixture (Figure D.30).

There are various ways of obtaining the required single enantiomer of a drug. For example, the synthesis of the racemic mixture may be carried out, followed by separation using chiral chromatography (normal

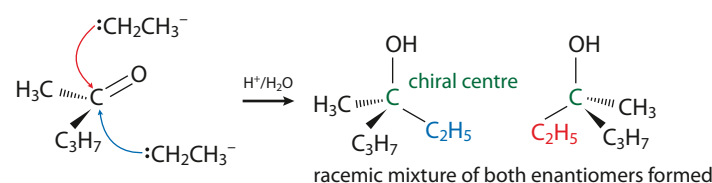


Figure D.30 A racemic mixture is formed because the reagent can attack from above or below.

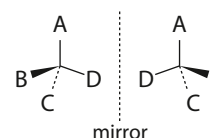
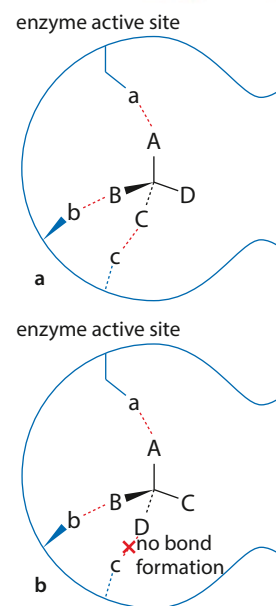


Figure D.27 Representation of a chiral drug binding to a theoretical enzyme active site: **a** one enantiomer can form three bonds with groups at the active site and is active; **b** the other enantiomer can only form two bonds with groups at the active site and is inactive.

chromatography does not separate enantiomers), or a synthetic reaction may be used that selectively produces one of the enantiomers of the product – this is known as **stereoselective (asymmetric) synthesis**.

One method of achieving stereoselective synthesis involves the use of a **chiral auxiliary**. This chiral auxiliary is a pure enantiomer and combines with the non-chiral reactant to form a chiral intermediate. The physical presence of the chiral auxiliary allows the reagent in the next stage of the synthesis to approach from one side of the molecule only, so forcing the reaction to follow a certain path that favours the production of one of the possible enantiomers (Figure D.31). Once the reaction is complete, the chiral auxiliary is removed to leave the desired enantiomer. The chiral auxiliary can then be recycled for use in other reactions.

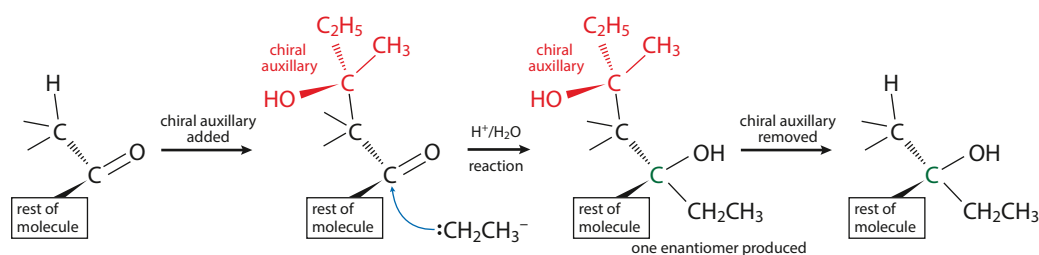


Figure D.31 A chiral auxiliary favours the formation of one enantiomer.

A chiral auxiliary is one enantiomer of an optically active substance that is temporarily incorporated into a non-chiral molecule to produce a single enantiomer of a product in an organic synthesis reaction.

Paclitaxel

One process in which chiral auxiliaries have been used successfully is in the semi-synthesis (from a natural precursor) of the anticancer drug paclitaxel (Taxol) (Figure D.32).

Paclitaxel is used to treat several forms of cancer – mainly breast, ovarian and lung cancer. It is usually given intravenously as part of a course of **chemotherapy** to treat cancer. Paclitaxel acts by preventing cell division – it does this by binding to microtubules in the cytoplasm, preventing them from breaking down during cell division.

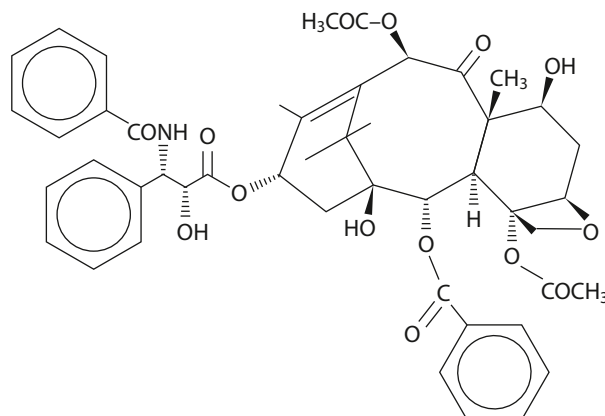


Figure D.32 The structure of paclitaxel.

Chemotherapy is the treatment of cancer with drugs. These drugs destroy cancer cells or stop them from dividing.

Paclitaxel was originally obtained from Pacific yew tree bark – however, it took the bark from more than one tree to provide enough paclitaxel to treat just one patient and so semi-synthetic processes were developed that involved making paclitaxel from another natural product derived from the needles of yew trees. Semi-synthesis of the drug allowed it to be made on a larger scale and reduced the environmental impact – extracting the drug from its natural source results in killing of the trees. Nowadays paclitaxel is also made by fermentation using plant cell cultures.

Using a polarimeter to distinguish between enantiomers

The enantiomers of an optically active substance can be distinguished using a polarimeter because they rotate the plane of plane-polarised light in opposite directions. A simple polarimeter consists of a source of light (usually a sodium lamp producing one specific wavelength), two polarising filters, a sample tube and a scale to measure the degree of rotation of the plane-polarised light (Figure D.33).

Polarimetry is usually carried out on samples in solution. First, the solvent in which the test substance is to be dissolved is put in the sample tube, then the polarising filters are rotated until the maximum amount of light passes through. At this point the two polarising filters are exactly parallel.

The solvent is then replaced by a solution of the sample in the same solvent and the polarising filter rotated again until the light coming through is of maximum brightness again. The direction in which the filter must be rotated and the angle through which the light is rotated are recorded.

An enantiomer that rotates plane-polarised light clockwise (to the right – dextrorotatory) is called the (+)-enantiomer and one that rotates the plane anticlockwise (to the left – levorotatory) is called the (–)-enantiomer.

Just which configuration (arrangement of groups around the chiral centre) of an enantiomer corresponds to the direction in which light is rotated can be worked out only by determining the absolute configuration using X-ray crystallography and the rotation of plane-polarised light using a polarimeter. We cannot just look at a particular enantiomer's three-dimensional structure and say that it rotates plane-polarised light to the right or to the left.

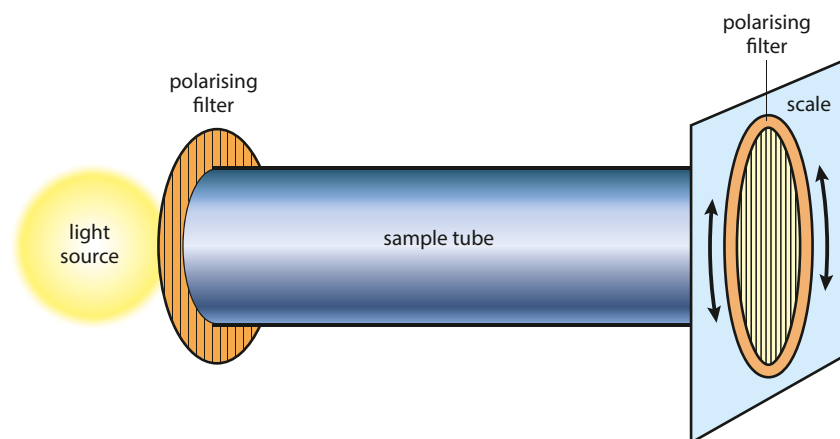


Figure D.33 A simple polarimeter.

An enantiomer can be identified from its specific rotation. The specific rotation, $[\alpha]$, is worked out from the angle through which the plane-polarised light is rotated in degrees (α), the path length of solution that the light passes through (l) in dm and the concentration of the solution (c) in g cm^{-3} :

$$[\alpha] = \frac{\alpha}{cl}$$

The specific rotation also depends on the wavelength of light used, the temperature (usually 20°C) and the solvent. The specific rotation of an enantiomer under a given set of conditions is characteristic of the compound and can be used to identify it by comparing with tables of literature values. However, this method does rely on having a pure sample of one enantiomer because contamination with the other enantiomer, or any other optically active compounds, will affect the angle through which the plane-polarised light is rotated.

Nature of science

Science is both systematic and creative. Systematic screening of a large number of plant extracts identified paclitaxel as a possible anticancer drug. When the demand for paclitaxel exceeded the supply from natural sources scientists had to develop ways of making this important drug. A great deal of creativity was involved in designing ways of making paclitaxel.

Learning objectives

- Understand what is meant by radioactive decay and balance equations involving α and β particles
- Calculate the amount of material remaining after a certain period of time when an isotope decays
- Understand what particles are emitted when nuclear medicine is used
- Understand that radiotherapy can be external or internal
- Understand the reasons for the use of certain radioisotopes
- Understand that targeted alpha therapy and boron-neutron-capture therapy are used in the treatment of cancer

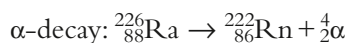
Beta particles are sometimes written as β^- to distinguish them from positrons, which are written β^+ .

D8 Nuclear medicine (HL)

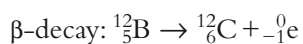
Radioactive decay processes

Radioactive decay involves changes in the nucleus of an atom resulting in particles and sometimes also electromagnetic radiation being emitted from the nucleus. The two main particles that are emitted are α particles (${}^4_2\text{He}^{2+}$, helium nuclei – two protons and two neutrons) and β particles (electrons formed when a neutron turns into a proton and an electron). The emission of γ rays (high-energy electromagnetic radiation) may also accompany radioactive decay. Other particles that can be emitted from nuclei include positrons, which are positively charged electrons produced when a proton is converted to a neutron in the nucleus.

Examples of radioactive decay processes are:



${}^4_2\alpha$ could also have been shown as ${}^4_2\text{He}$.



${}^0_{-1}\text{e}$ could also have been shown as ${}^0_{-1}\beta$.

Equations for nuclear decay must balance in terms of the total mass number and the total atomic number on each side.

When a nucleus undergoes α decay, the mass number will go down by four and the atomic number will go down by two. The element formed will be two places to the left in the periodic table.

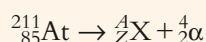


Because the emission of a β particle results from a neutron in the nucleus becoming a proton, the mass number does not change but the atomic number goes up by one (the atomic number is the number of protons). The element formed will be one place to the right in the periodic table.

Worked examples

D.9 Astatine-211 can undergo α decay by emitting α particles. Determine the identity of the isotope formed.

The equation for the reaction is:



The total mass number on the right-hand side must equal 211; therefore $A + 4 = 211$ and $A = 207$

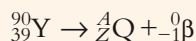
The total atomic number on the right-hand side must equal 85; therefore $Z + 2 = 85$ and $Z = 83$

The atomic number defines the element and the element with atomic number 83 is bismuth.

The isotope formed is ${}_{83}^{207}\text{Bi}$.

D.10 Yttrium-90 can undergo β decay by emitting β particles. Determine the identity of the isotope formed.

The equation for the reaction is:



The total mass number on the right-hand side must equal 90; therefore $A + 0 = 90$ and $A = 90$

The total atomic number on the right-hand side must equal 39; therefore $Z - 1 = 39$ and $Z = 40$

The atomic number defines the element and the element with atomic number 40 is zirconium.

The isotope formed is ${}_{40}^{90}\text{Zr}$.

? Test yourself

5 Determine the identity of the isotope formed in each of the following decay processes:

- a** indium-115 undergoes β decay
- b** radium-224 undergoes α decay
- c** nickel-63 undergoes β decay
- d** thorium-229 undergoes α decay

6 Work out whether each of the following processes involves emission of an α particle or a β particle:

- a** ${}_{71}^{177}\text{Lu} \rightarrow {}_{72}^{177}\text{Hf} + ?$
- b** ${}_{79}^{173}\text{Au} \rightarrow {}_{77}^{169}\text{Ir} + ?$
- c** ${}^{66}\text{Cu} \rightarrow {}^{66}\text{Zn} + ?$
- d** ${}^{104}\text{Mo} \rightarrow {}^{104}\text{Tc} + ?$
- e** ${}^{278}\text{Rg} \rightarrow {}^{274}\text{Mt} + ?$

Half-life

Radioactive decay is a random process and it is impossible to predict when any one nucleus will decay, but on average a sample of a particular radioisotope containing a very large number of atoms will decay with a constant half-life.

Half-life is the time it takes for the number of radioactive nuclei present in a sample at any given time to fall to half its value.

Half-life ($t_{1/2}$) varies from isotope to isotope – for example, the half-life of ^{226}Ra is 1600 years but that of ^{224}Ra is 3.7 days. Half-life is independent of the mass of a radioactive sample – the half-life is the same whether 1 g or 1 kg of a particular isotope is present.

Figure D.34 shows a graph of the decay of an isotope with half-life 2 s. Every 2 s the number of nuclei remaining decreases by half.

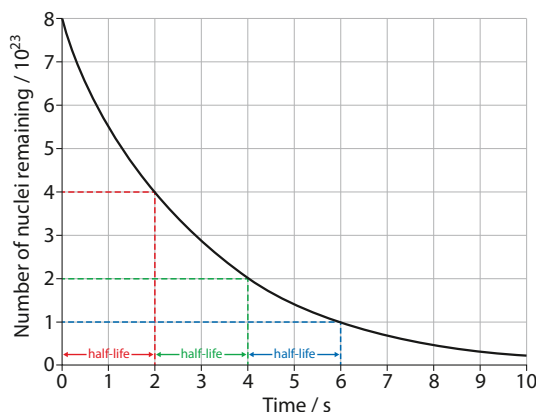


Figure D.34 The decay of a radioisotope with half-life 2 s – this is exponential decay.

Calculations involving whole numbers of half-lives

Radium-226 is an α -emitter with a half-life of approximately 1600 years – so if we start with 1 g of pure ^{226}Ra , after 1600 years there will be 0.5 g present, after a further 1600 years there will be 0.25 g present and after a total of 4800 years (three half-lives) there will only be 0.125 g of radium-226 left:

$$1 \text{ g} \xrightarrow[1600 \text{ years}]{\text{half-life}} 0.5 \text{ g} \xrightarrow[1600 \text{ years}]{\text{half-life}} 0.25 \text{ g} \xrightarrow[1600 \text{ years}]{\text{half-life}} 0.125 \text{ g}$$

Half-life may also be expressed in terms of the activity of a sample (the number of nuclei that decay per second). The half-life is then the time taken for the activity to drop to half its original value.

Worked examples

D.11 Germanium-71 has a half-life of 11 days. If there were originally 2.00 mg of this isotope present in a sample, calculate the mass remaining after 44 days.

It takes 11 days for the amount to fall by half, therefore there will be 1.00 mg present after 11 days. This will drop to 0.500 mg after another 11 days, 0.250 mg after another 11 days and 0.125 mg after a further 11 days:

$$2.00 \text{ mg} \xrightarrow{11 \text{ days}} 1.00 \text{ mg} \xrightarrow{11 \text{ days}} 0.500 \text{ mg} \xrightarrow{11 \text{ days}} 0.250 \text{ mg} \xrightarrow{11 \text{ days}} 0.125 \text{ mg}$$

Therefore the mass of germanium-71 remaining will be 0.125 mg.

D.12 The half-life of uranium-238 is 4.5×10^9 years. Calculate how long it would take 32 g of uranium-238 to decay to 1 g.

This decay involves five half-lives:

$$32 \text{ g} \xrightarrow{\text{half-life}} 16 \text{ g} \xrightarrow{\text{half-life}} 8 \text{ g} \xrightarrow{\text{half-life}} 4 \text{ g} \xrightarrow{\text{half-life}} 2 \text{ g} \xrightarrow{\text{half-life}} 1 \text{ g}$$

So the total time is $5 \times 4.5 \times 10^9 = 2.3 \times 10^{10}$ years.

D.13 Calculate the half-life of protoactinium-233 if it takes 108 days for 100 mg of the element to decay to 6.25 mg.

This decay will take four half-lives:

$$100 \text{ mg} \xrightarrow{\text{half-life}} 50 \text{ mg} \xrightarrow{\text{half-life}} 25 \text{ mg} \xrightarrow{\text{half-life}} 12.5 \text{ mg} \xrightarrow{\text{half-life}} 6.25 \text{ mg}$$

So 108 days is equivalent to four half-lives. One half-life is therefore $\frac{108}{4} = 27$ days.

D.14 Calculate the time taken for the activity of a sample of rubidium-83 to fall to 12.5% of its original value given that its half-life is 86 days.

The data here are given in terms of activity instead of mass, but the answer is worked out in the same way:

$$100\% \xrightarrow{86 \text{ days}} 50\% \xrightarrow{86 \text{ days}} 25\% \xrightarrow{86 \text{ days}} 12.5\%$$

Three half-lives are required for this decay and so the total time is $3 \times 86 = 258$ days.

? Test yourself

- 7 In each of the following questions, calculate the amount remaining from a 100 mg sample of the given radioisotope after the specified time.
- ^{105}Rh has a half-life of 35 h. Calculate the mass remaining after 70 h.
 - ^{209}Po has a half-life of 105 y. Calculate the mass remaining after 420 y.
 - ^{219}Rn has a half-life of 3.96 s. Calculate the mass remaining after 39.6 s.
- 8 Calculate the half-lives of each of the following radioisotopes.
- It takes 180 days for 80 mg of iron-59 to decay to 5 mg.
 - It takes 2.1×10^{12} years for 60 mg of platinum-190 to decay to 7.5 mg.
- 9 Calculate how long it will take for each of the following decay processes.
- 32 mg of silicon-32 (half-life 160 y) to decay to 1 mg.
 - 56 mg of mendelevium-258 (half-life 56 d) to decay to 7 mg.

Calculations involving non-integral numbers of half-lives

Radioactive decay is a first-order process, so the rate of decay is proportional to the number of undecayed nuclei remaining. We normally discuss the rate of decay in terms of the activity (A), which is the number of nuclei which decay per second, and so we can write a rate equation for radioactive decay as:

$$A = \lambda N$$

where λ is the decay constant and N is the number of undecayed nuclei present. The unit of activity is the becquerel (Bq), which is equivalent to one disintegration (decay) per second.

This can be compared to a first-order rate equation for a chemical reaction:

$$\text{rate} = k[\text{X}]$$

as described in Topic 6. It can be seen that the decay constant is equivalent to a rate constant.

Integration of the rate equation for radioactive decay in calculus form produces the equation:

$$N = N_0 e^{-\lambda t}$$

where N_0 is the initial number of undecayed nuclei present and N is the number of undecayed nuclei present at time t .

Further mathematical manipulation produces an equation that relates the decay constant to the half-life:

$$\lambda = \frac{\ln 2}{t_{\frac{1}{2}}}$$

It can be seen from these equations that, like the first-order rate constant, the units of the activity constant are time^{-1} . If the activity is measured in becquerels then the activity constant should have units of s^{-1} .

The rate equation for radioactive decay using calculus notation is:

$$\frac{dN}{dt} = -\lambda N$$

This equation could also be written in the form

$$\ln\left(\frac{N}{N_0}\right) = -\lambda t$$

Worked examples

D.15 The half-life of rutherfordium-104 is 65 s. Calculate the decay constant and the percentage of a 1.00 μg sample remaining after 3.00 minutes.

$$\begin{aligned}\lambda &= \frac{\ln 2}{t_{\frac{1}{2}}} \\ &= \frac{0.693}{65} \\ &= 0.011 \text{ s}^{-1}\end{aligned}$$

This answer has been rounded to two significant figures but more figures have been carried through for further calculations.

Next we need to use $N = N_0 e^{-\lambda t}$. Both λ and t must be in the same units of time – so if λ is in s^{-1} , t must be in seconds. 3.00 minutes is 180 s.

We are trying to find the ratio $\frac{N}{N_0}$, and we have:

$$\begin{aligned}\frac{N}{N_0} &= e^{-\lambda t} \\ &= e^{-0.011 \times 180} \\ &= 0.147\end{aligned}$$

This is multiplied by 100 to get a percentage and so about 15% remains undecayed after 3.00 minutes.

D.16 Given that the activity of a 1.00 μg sample of nitrogen-13 is $5.35 \times 10^{13} \text{ Bq}$, calculate the half-life and the mass left after 45 minutes. The mass of a nitrogen-13 atom is $2.16 \times 10^{-23} \text{ g}$.

The number of nitrogen-13 atoms in 1.00 μg is:

$$\begin{aligned}\frac{1.00 \times 10^{-6}}{2.16 \times 10^{-23}} \\ \text{i.e. } 4.63 \times 10^{16} \text{ atoms}\end{aligned}$$

$A = \lambda N$ so:

$$\begin{aligned}\lambda &= \frac{A}{N} \\ &= \frac{5.35 \times 10^{13}}{4.63 \times 10^{16}} \\ &= 1.16 \times 10^{-3} \text{ s}^{-1}\end{aligned}$$

$$\begin{aligned}t_{\frac{1}{2}} &= \frac{\ln 2}{\lambda} \\ &= \frac{0.693}{1.16 \times 10^{-3}} \\ &= 600 \text{ s}\end{aligned}$$

To calculate the mass left after 45 minutes we use:

$$N = N_0 e^{-\lambda t} \text{ in the form } \frac{N}{N_0} = e^{-\lambda t}$$

The decay constant is in units of s^{-1} so we must convert the time to seconds:

$$t = 45 \times 60 = 2700 \text{ s}$$

$$\begin{aligned}\frac{N}{N_0} &= e^{-\lambda t} \\ &= e^{-1.16 \times 10^{-3} \times 2700} \\ &= 0.0442\end{aligned}$$

This is the proportion of the sample *undecayed* after 45 minutes, so the mass of nitrogen-13 left at the end is 0.0442 times the original mass:

$$\begin{aligned}\text{mass} &= 0.0442 \times 1.00 \\ &= 0.0442 \mu\text{g, or } 4.42 \times 10^{-8} \text{ g}\end{aligned}$$

It is important to be consistent with units in these questions. The decay constant in the above example could also have been written as 0.0693 min^{-1} . This is just 60 times the decay constant given.

Using the decay constant in this form means that time can be used in minutes. Substituting in $\frac{N}{N_0} = e^{-\lambda t}$ we get:

$$\frac{N}{N_0} = e^{-0.0693 \times 45}, \text{ which gives the same answer as above.}$$

D.17 1.00 kg of a particular rock was believed to contain 0.120 g of potassium-40 when it was originally formed. Analysis of the rock has determined that it now contains 0.0500 g of potassium-40. Calculate the age of the rock given that the half-life of potassium-40 is 1.25×10^9 years.

We will need to use $\frac{N}{N_0} = e^{-\lambda t}$ and so must first work out the decay constant:

$$\begin{aligned}\lambda &= \frac{\ln 2}{t_{1/2}} \\ &= \frac{0.693}{1.25 \times 10^9} \\ &= 5.54 \times 10^{-10} \text{ y}^{-1}\end{aligned}$$

We can use mass instead of number of atoms in the following equation because any conversion factor will simply cancel out in the ratio.

$$\begin{aligned}\frac{N}{N_0} &= e^{-\lambda t} \\ \frac{0.0500}{0.120} &= e^{-5.54 \times 10^{-10} t}\end{aligned}$$

$$\text{So } e^{-5.54 \times 10^{-10} t} = 0.4167$$

Taking the natural log of both sides:

$$\begin{aligned}-5.54 \times 10^{-10} t &= \ln 0.4167 \\ -5.54 \times 10^{-10} t &= -0.875\end{aligned}$$

Exam tip

$\frac{N}{N_0} = e^{-\lambda t}$ can be used more conveniently in the form $\ln\left(\frac{N}{N_0}\right) = -\lambda t$, which makes the mathematical manipulation simpler in this type of problem.

$$\text{So } t = \frac{0.875}{5.54 \times 10^{-10}}$$

$$= 1.58 \times 10^9 \text{ y}$$

The decay constant was expressed in y^{-1} so the time will come out in years.

Exam tip

As a final check we can consider if this answer seems reasonable. The half-life of potassium-40 is 1.25×10^9 years, therefore 0.120 g should decay to 0.0600 g in this time. We are trying to find the time that it takes to decay to 0.0500 g and therefore would expect it to be slightly longer than one half-life, which our answer is.

? Test yourself

- 10 Calculate the decay constant for each of the following:
 - a the half-life of nobelium-259 is 58 minutes
 - b the half-life of rubidium-83 is 86.2 days
 - c the half-life of neodymium-144 is 2.1×10^{15} years
- 11 Assuming that you start in each case with $5.00 \mu\text{g}$ of the isotopes in question 10, calculate how long it will take for the amount of the isotope to drop to $1.00 \mu\text{g}$.
- 12 Calculate what percentage of the stated isotope is left after the stated time.
 - a carbon-14 (half-life 5730 years) after 10 000 years
 - b iodine-131 (half-life 8.04 days) after 3 weeks
- 13 Given that the activity of a $2.00 \mu\text{g}$ sample of osmium-191 is $3.27 \times 10^9 \text{ Bq}$, calculate its half-life and the mass left after 6 weeks. The mass of an osmium-191 atom is $3.19 \times 10^{-22} \text{ g}$.

Radioactivity in medicine

Alpha, beta, gamma, proton, neutron and positron emissions are used in nuclear medicine.

Radioactivity is used in the diagnosis and treatment of disease.

In diagnostic applications, radioactive atoms are incorporated in pharmaceutical molecules or biochemical molecules (such as hormones) and injected into the body. These molecules travel round the body and their progress and interaction with cells and organs can be monitored using a detector that picks up the radiation emitted. Radioisotopes commonly used in imaging are gamma and/or positron (β^+) emitters, such as technetium-99m (γ) and fluorine-18 (β^+).

Radiotherapy (radiation therapy) refers to the treatment of a disease, usually cancer, using radiation. Radioisotopes for radiotherapy commonly emit α particles, β particles and γ rays. Proton-beam therapy, using protons from a particle accelerator, has also been used to treat some cancers.

A related technique is neutron therapy, where a beam of protons from a particle accelerator strikes a beryllium target to produce a beam of neutrons which can be effective against tumours. Positron emitters are commonly used in the diagnosis of cancer but there has been some interest in using them in therapy as well.

Radiotherapy

α , β and γ rays are called *ionising radiation* because they cause the formation of ions (by ejection of electrons) when they interact with matter. Ionising radiation can damage cells and the main effect comes from damage caused to DNA.

Ionising radiation can either interact directly with the DNA, causing ionisation and a change of structure, or there can be indirect effects due to the formation of free radicals from other species such as water. The most common substance in our body is water and when ionising radiation interacts with water molecules they can become ionised. The ion and electron generated can go on to react further to produce free radicals such as the hydroxyl radical ($\text{HO}\cdot$). Of the free radicals generated in these processes, the hydroxyl radical is probably the most dangerous – when it interacts with DNA it can trigger a series of reactions which results in damage to the DNA. Free radicals can also cause damage to proteins (enzymes) and lipids in cells.

Most cells in the body divide and replicate themselves – for example, to replace dead cells – and this happens in a controlled manner. Cancer involves cells that have been changed in some way(s) so that they replicate continuously – this happens in an uncontrolled manner. Because radiation mainly affects DNA, it has the greatest effect on cells that are replicating (using DNA), that is cancer cells. Healthy cells, which are not replicating, are more able to repair any damaged DNA and recover from the effect of the radiation. However, there are some cells in the body that do replicate more frequently and these include cells in hair follicles, red blood cells and tissue that is growing (as in children). These cells are also affected more by radiation and this explains why the side effects of radiotherapy can include hair loss and anemia.

The aim of radiotherapy is to kill cancer cells using radiation – but cause the minimum amount of damage to surrounding tissue and other cells in the body.

Radiotherapy can involve an external or internal source of radiation.

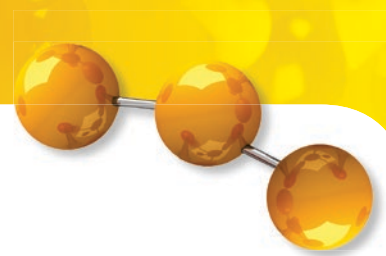


Figure D.35 A person undergoing radiotherapy using an external source of radiation.

External radiotherapy involves targeting radiation from a machine that generates a beam of radiation onto a specific area of the body (Figure D.35). Different machines produce beams of γ rays, (e.g. from cobalt-60), protons, electrons or X-rays.

There are two forms of treatment involving internal sources of radiation:

- brachytherapy – this involves putting a solid source of radioactivity into or near the tumour within the body. This is used to treat several types of cancer including prostate cancer and cancers of the head, neck, womb or cervix. Radioisotopes used in brachytherapy include palladium-103 (γ -emitter) and cobalt-60 (γ -emitter). Implants may be temporary (inserted and then removed later) or permanent.
- radioisotope therapy – using a liquid that is injected intravenously or taken orally. For instance, a patient with thyroid cancer may be treated



with iodine-131 (β -emitter) by being given a capsule or solution containing radioactive iodine-131 (as sodium iodide) to take orally. The iodine is taken up by cancerous cells in the thyroid gland and the radiation kills them (and also healthy cells).

Side effects of radiotherapy

Radiotherapy does not just affect cancer cells but also damages healthy cells in the area of treatment.

Common side effects of radiotherapy include:

- hair loss – this can occur where the beam enters the body and where it leaves the body; it is usually a temporary effect
- nausea – this is most likely when the treatment area is near the stomach; it is a temporary effect
- fatigue – tiredness can be caused by anemia due to red blood cells being destroyed during the treatment; a temporary effect
- sterility – this can occur if the treatment area includes the ovaries and testes; a permanent effect.

Radioisotopes used in nuclear medicine

Technetium-99m is the most common radioisotope used in medicine.

Technetium-99m is used widely as a radioactive tracer in medical imaging to diagnose illnesses. A radioactive tracer is introduced into the body and taken up by cells. It emits γ -radiation (or positrons, which produce γ -rays) that passes out through the body, and is detected using a γ -camera, which is designed to pick up such radiation.

There are several reasons why technetium-99m is especially suitable for this use:

- It is very close to being a pure emitter of γ -rays – γ -rays can pass through the body but α and β particles are not sufficiently penetrating. $^{99\text{m}}\text{Tc}$ decays by a process called ‘isomeric transition’ – the ‘m’ in ‘99m’ means that its nuclei are in an metastable state – a reasonably long-lived excited nuclear state – and it can decay to ^{99}Tc by giving out this excess energy in the form of γ -ray photons. The energy of the γ -rays produced is similar to that of X-rays used in X-ray machines – the γ -rays are penetrating enough to pass out through the body, but not so penetrating that they are excessively dangerous.
- It has a half-life of six hours – this is long enough to allow it to travel round the body, but short enough that the patient does not remain radioactive for long.
- Technetium has a reasonably extensive chemistry and so its radioactive atoms can be incorporated into compounds that are soluble and can be transported round the body. Radioactive $^{99\text{m}}\text{Tc}$ can be made into the technetate(VII) ion – (TcO_4^-) , which is water-soluble but not as strong an oxidising agent as the manganate(VII) ion (Mn is in the same group as Tc).

Internal radiotherapy procedures usually use β -emitters rather than γ -emitters – β particles interact with matter more effectively than γ -rays and lose all their energy within a few millimetres in the body, causing a great deal of damage to cells in a localised area but little damage to cells

outside that area. γ -rays are able to pass out of the body without losing much energy and so γ -emitters are more useful for imaging purposes. Gamma rays pass through more tissue and cause damage to cells over a greater area but less damage to individual cells.

Yttrium-90 and lutetium-177 are used in radiotherapy – they are both β -emitters with relatively short half-lives.

Yttrium-90 is a β -emitter with a half-life of 64 hours and is used for the treatment of liver cancer. In the treatment, tiny beads (about $30\mu\text{m}$ in diameter) containing ^{90}Y are injected into the artery carrying blood to the liver and act locally on the tumour there, killing cells within a very short range.

Lutetium-177 is a β -emitter and γ -emitter with a half-life of 6.71 days. It is used for targeted radiotherapy by being incorporated into molecules that can bind to receptors on certain types of cell. The radiation then destroys only a particular type of cells within a very limited area. ^{177}Lu can be used for treatment of neuroendocrine cancers – the neuroendocrine system is the system of nerves and glands that has the role of producing hormones in the body. The fact that ^{177}Lu is a γ -emitter as well as a β -emitter means that it can also be used for imaging purposes.

Targeted alpha therapy

Alpha particles are relatively large and highly charged and cause a great deal of damage to cells in a very small area. They are not very penetrating and so cannot be used as an external source of radioactivity – they are not able to penetrate skin.

Targeted alpha therapy (TAT) is a relatively new form of treatment that is still very much in the research stage. It involves using some way of bringing the source of α particles specifically to the cancer cells. The range of α particles in the body is typically $50\text{--}100\mu\text{m}$ (about the size of a human cell) and because they lose all their energy in a very small space they cause a great deal of damage to cells (more than β particles).

TAT can be designed to attack, as far as possible, just cancer cells by using monoclonal antibodies – antibodies that are all the same shape. Antibodies are proteins of the immune system that bind to specific receptors on certain cells (or foreign particles) to target them for destruction in the body. Monoclonal antibodies can be made to target a specific type of cancer cell (bearing a specific receptor on its surface) and can be labelled with an α -emitting radioisotope. The antibodies travel through the body and attach to just this one type of cell, carrying the radioisotope with them. Decay of the radioisotopes produces α particles, which destroy the cancerous cell. Up to about 20 α particles may be required to kill one cell. The antibodies will target one particular type of cell anywhere in the body and so TAT has the potential to treat cancers that have spread throughout the body.

Various radioisotopes have been suggested for this type of radiotherapy, including astatine-211 (half-life 7.2 hours) and lead-212 (half-life 10.6 hours)

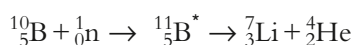
A different type of TAT involves the use of radioactive radium chloride (radium-223, an α - and γ -emitter with a half-life of 11.4 days) to treat certain cancers that have spread to the bones. Radium is in the same group in the periodic table as calcium, which makes up a large proportion of bones, and is treated in the same way as calcium in the body. After

injection, the radium travels through the bloodstream to the bones, where it is taken up by cancerous cells in the bone (these are dividing more rapidly and will need a bigger supply of calcium/radium) and destroys the cells in the immediate vicinity (i.e. the cancerous cells) but healthy cells receive only small amounts of radiation or none at all. In this way it can target areas of cancer throughout the bones in the body.

Boron-neutron-capture therapy

Boron-neutron-capture therapy (BNCT) has the potential to be a promising form of radiotherapy for the treatment of head and neck cancers.

BNCT relies on the fact that when non-radioactive boron-10 atoms, which have been taken up by cancer cells, are irradiated with a neutron beam from outside the body, they can capture neutrons to produce a high-energy form of boron-11, which can undergo fission to produce α particles and lithium nuclei:



These massive, charged particles have a very short path length in the body and get rid of their energy in a very small space causing a great deal of damage to cells.

An essential part of BNCT is to make sure that tumour cells take up sufficient amounts of boron-10 and various drugs have been developed to achieve this.

Magnetic resonance imaging

Magnetic resonance imaging (MRI) involves the use of nuclear magnetic resonance (NMR) to produce three-dimensional images of the internal organs. Although the word 'nuclear' is in the name, the process involved in this technique is very different to those described above. NMR involves no changes to the nucleus of atoms but rather involves the change in orientation of a spinning nucleus relative to an external magnetic field.



Figure D.36 An MRI scanner.



Figure D.37 A coloured MRI scan of a whole human body.

An MRI body scanner (Figure D.36) is an NMR spectrometer in which a patient can be placed. The scanning takes 15–45 minutes, and the patient is required to lie still for this length of time. MRI interacts with the protons in water molecules (and other molecules such as fat) in cells in the body. Water molecules in cells in different organs are in slightly different environments, so the various organs in the body can be differentiated.

MRI is a safe, non-invasive technique for scanning organs in the body, and when the data are analysed using a computer, it is possible to obtain a three-dimensional scan of the body (Figure D.37). The only radiation involved is that in the radiofrequency part of the electromagnetic spectrum – side effects are rare and very minor.

Nature of science

In many medical processes, such as the use of radioactivity, scientists must consider whether the benefits outweigh the risks of the procedure. There will be many factors involved in such an analysis and an objective approach is essential.

Learning objectives

- Understand that, after synthesis, a drug must be extracted from the reaction mixture and purified
- Describe the processes of extraction and purification
- Understand some factors that affect the solubility of organic compounds
- Understand what is meant by Raoult's law
- Understand the use of fractional distillation in the purification of an organic compound
- Understand that a combination of spectroscopic techniques can be used to identify an organic compound
- Understand that steroids can be detected in urine samples using gas chromatography–mass spectrometry
- Understand that alcohol can be detected in the breath using a breathalyser

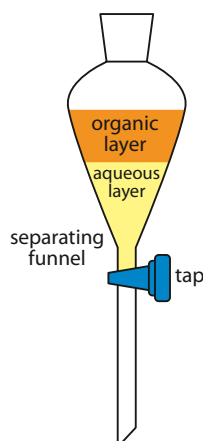
D9 Drug detection and analysis (HL)

Extraction of drugs

After making a drug, the next thing to do is extract it from the reaction mixture – this is often called the work-up by organic chemists and usually relies on the solubility of the product varying from one solvent to another.

The first step in extraction often involves adding the reaction mixture to ice-water. If the product is a solid and insoluble in water it might precipitate at this stage and can be removed from the reaction mixture by filtration.

If the product does not precipitate out then it can be extracted from the aqueous mixture by solvent extraction. This involves shaking the aqueous mixture with a solvent that is immiscible with water in a separating funnel – common solvents include ethoxyethane and ethyl ethanoate. A separating funnel is shown in Figure D.38.



D.38 A separating funnel.

The separating funnel is shaken and the various components of the mixture partition between the two solvents according to their solubility in each – the pharmaceutical is likely to be an organic, non-polar compound that will dissolve much better in the organic solvent. Any inorganic, polar substances will dissolve much better in the aqueous layer. The mixture is allowed to settle and two layers should form – an organic layer and an aqueous layer. The tap is opened and the layers separated.

Better separation is obtained by using several small portions of organic solvent rather than one large one. The samples of organic solution are then combined and a drying agent (such as anhydrous magnesium sulfate) added to remove any water. The drying agent is filtered off and the organic solvent removed, usually by using a rotary evaporator (Figure D.39) which takes the solvent off under reduced pressure – this means that the solution does not have to be heated so much which is important if the product is thermally unstable.

Solubility

In order to identify suitable solvents for extraction, it is necessary to understand the factors that affect the solubility of organic compounds. The general rule is 'like dissolves like', which means that molecules will tend to dissolve in solvents with similar intermolecular forces – polar molecules will tend to dissolve in polar solvents such as water; and non-polar substances will dissolve in non-polar, organic solvents and generally be insoluble in water.

The ability to form hydrogen bonds often makes substances more soluble in water. As a general rule, small, highly polar molecules that are able to participate in hydrogen bonding will be soluble in water but larger, less polar molecules will not be.

We can look at some factors that affect solubility by considering some examples.

The structures of ethanoic acid and methyl methanoate, which are isomers of each other, are shown in Figure D.40. Both molecules are fairly small and polar but ethanoic acid has an H attached directly to an O and is therefore able to hydrogen bond to water, making it more soluble in water.

The structure of ethyl ethanoate is shown in Figure D.41. This is also polar but has more non-polar regions (CH₃ groups) than methyl methanoate and is therefore considerably less soluble in water.

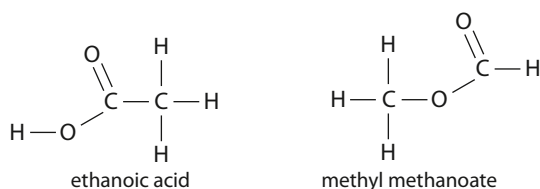


Figure D.40 The structures of ethanoic acid and methyl methanoate.

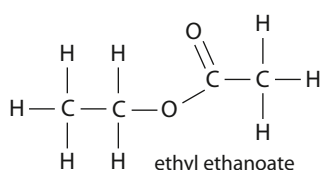


Figure D.41 The structure of ethyl ethanoate.



Figure D.39 A rotary evaporator.

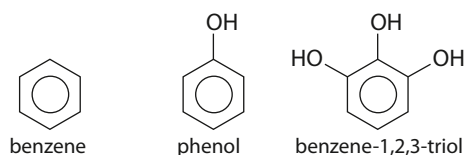


Figure D.42 The structures of benzene, phenol and benzene-1,2,3-triol.

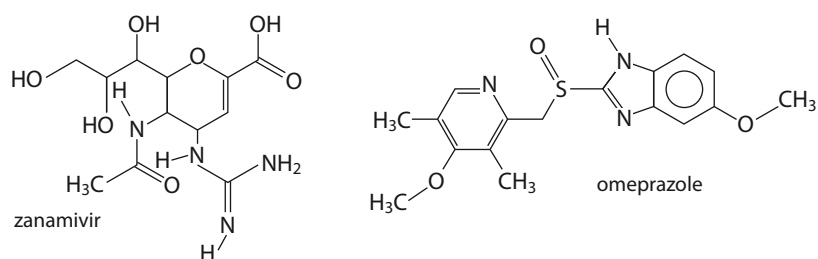


Figure D.43 The structures of zanamivir and omeprazole.

Generally, the presence of more OH groups or NH_2/NH groups in a molecule increases water solubility – phenol is about 45 times more soluble in water than benzene is, and benzene-1,2,3-triol is about five times more soluble than phenol (Figure D.42). The OH and NH/ NH_2 groups are able to form hydrogen bonds to water.

Zanamivir (Figure D.43), with lots of polar groups, is soluble in water, but omeprazole is about 70 times less soluble and is regarded as being only slightly soluble in water.

Substances containing ions tend to be more soluble in water because of the formation of ion–dipole interactions. Phenol is therefore more soluble in sodium hydroxide solution than in water because it forms the phenoxide ion ($\text{C}_6\text{H}_5\text{O}^-$). Similarly, aspirin is less soluble in acidic solutions than in alkaline solutions because of the formation of ions in alkaline solution – the conversion of aspirin to its sodium salt to increase its solubility was discussed on page 15. Oseltamivir tablets (see page 37) contain the phosphate salt, which is much more soluble in water.

Purification

Once the organic compound has been extracted from the reaction mixture it must be purified. There are several methods for purifying compounds but the most common are recrystallisation, distillation/fractional distillation and chromatography.

Recrystallisation of solid compounds has already been described on page 13 and here we will concentrate on fractional distillation. This can be used to separate liquids that have quite similar boiling points – whereas simple distillation is used to separate liquids with very different boiling points or for separating a liquid from a non-volatile residue. The experimental set-up for fractional distillation is shown in Figure D.44. The column is made of moulded glass or is packed with small beads to give a very large surface area. The mixture of liquids is heated and the liquid with the lower boiling point is collected by condensing the vapour.

To understand how fractional distillation works you need to understand a bit more about liquid–vapour equilibria.

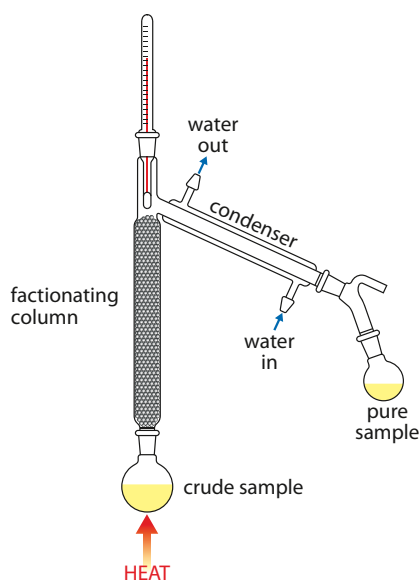


Figure D.44 The experimental set-up for fractional distillation.

Raoult's law

Phase equilibrium was introduced in Topic 7. The vapour above a liquid in a closed container exerts a pressure on the walls of the container. If the vapour is in equilibrium with the liquid, this pressure is called the vapour pressure (or 'equilibrium vapour pressure' or 'saturated vapour pressure') (Figure D.45).

Vapour pressure is the pressure exerted by a vapour in equilibrium with a liquid (or a solid).

For a pure liquid, the vapour pressure depends only on the nature of the liquid and the temperature. For a mixture of liquids, the vapour pressure also depends on how much of each liquid is present. Raoult's law states that the partial vapour pressure of any volatile component of an ideal solution is equal to the vapour pressure of the pure liquid multiplied by the mole fraction of that liquid in the solution. To understand what this means we need to introduce a few terms:

- *Volatile* – something that evaporates readily.
- *Ideal solution* – a mixture of liquids in which the intermolecular forces are the same as in the pure liquids – i.e. the tendency for a liquid to evaporate is the same in the pure liquid as in the solution. An example of a fairly-close-to-ideal solution is a mixture of hexane (C_6H_{14}) and heptane (C_7H_{16}).
- The *mole fraction* of a component, A, in a mixture is given by:

$$\text{mole fraction of A} = \frac{\text{number of moles of A}}{\text{total number of moles in the mixture}}$$

Mole fraction is given the symbol X , so in a mixture containing n_A moles of A and n_B moles of B, the mole fraction of A is given by:

$$X_A = \frac{n_A}{n_A + n_B}$$

Note that mole fraction is a ratio and therefore has no units.

- *Partial pressure* – refers to the pressure exerted by a particular gas in a mixture of gases. If the gases behave ideally, the partial pressure is the same as the pressure that the same amount of that particular gas would exert if it were in the container by itself. If the pressure exerted by a mixture of 80% nitrogen and 20% oxygen is 100 kPa, the partial pressure of nitrogen is 80 kPa and that of oxygen is 20 kPa. Partial pressure is calculated from:

$$\text{partial pressure of A} = \text{mole fraction of A} \times \text{total pressure}$$

or

$$P_A = X_A \times P_{\text{tot}}$$

Dalton's law of partial pressure states that for a mixture of ideal gases the total pressure is equal to the sum of the partial pressures:

$$P_{\text{tot}} = P_A + P_B + P_C + \dots$$

Returning to Raoult's law, it basically says that the contribution of each component of a mixture to the total vapour pressure depends on what the vapour pressure of the pure liquid is and how much is present in the mixture. It can be written:

$$P_A = X_A \times P_A^0$$

where P_A is the partial vapour pressure of A, X_A is the mole fraction of A in the mixture and P_A^0 is the vapour pressure of pure A.

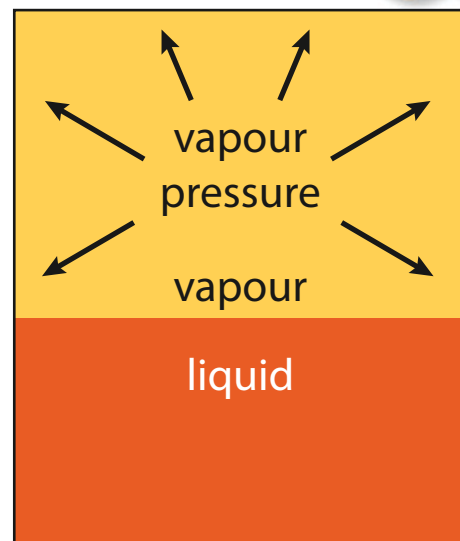


Figure D.45 A vapour in equilibrium with its liquid exerts a vapour pressure.

The sum of the mole fractions of the various components in a mixture is 1.

Note: 'mole fraction of A' here refers to the mole fraction in the vapour.

Figure D.46 shows how the vapour pressure of an ideal mixture varies with the composition of the mixture at constant temperature. At any point, the total vapour pressure is the sum of the partial vapour pressures of the components.

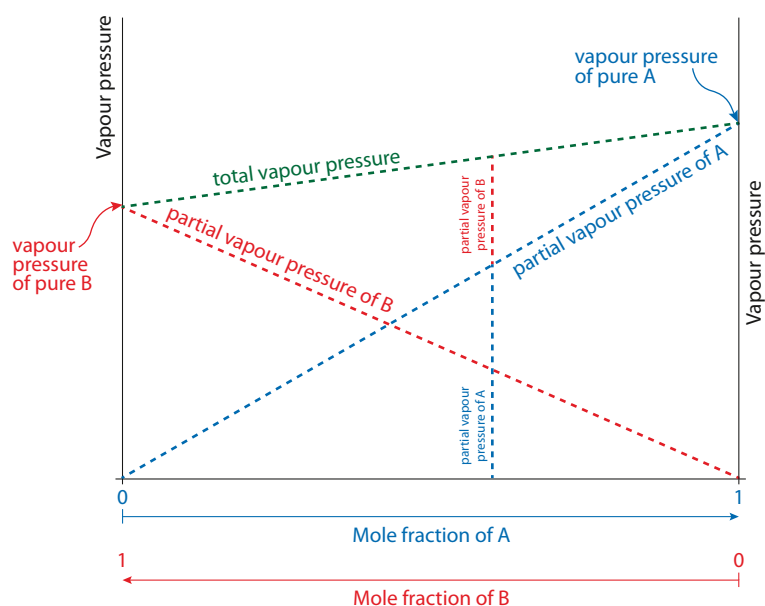


Figure D.46 The variation of vapour pressure with composition for an ideal solution.

Worked example

D.18 Given that the vapour pressures of pure hexane and heptane at 25 °C are 20.5 kPa and 6.10 kPa respectively, calculate the vapour pressure of an ideal solution containing 4.00 mol hexane and 4.00 mol heptane at 25 °C. Then calculate the mole fraction of each in the vapour.

Using Raoult's Law:

$$P_A = X_A \times P_A^0$$

$$X_A = \frac{n_A}{n_A + n_B}$$

$$X_{\text{hexane}} = \frac{4.00}{4.00 + 4.00} = 0.500$$

$$X_{\text{heptane}} = \frac{4.00}{4.00 + 4.00} = 0.500$$

$$P_{\text{hexane}} = 0.500 \times 20.5 = 10.25 \text{ kPa}$$

$$P_{\text{heptane}} = 0.500 \times 6.10 = 3.05 \text{ kPa}$$

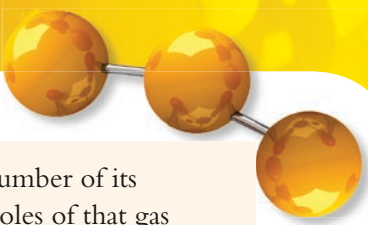
Using Dalton's Law of partial pressures:

$$P_{\text{tot}} = P_{\text{hexane}} + P_{\text{heptane}}$$

$$= 10.25 + 3.05$$

$$= 13.3 \text{ kPa}$$

So the total vapour pressure of the solution is 13.3 kPa.



If we assume that the gases behave ideally, the pressure exerted by a gas is proportional to the number of its molecules present, therefore the partial pressure of each gas is proportional to the number of moles of that gas present in the vapour phase. So, the mole fraction of hexane in the vapour phase is given by:

$$X_{\text{hexane}}^{\text{vap}} = \frac{P_{\text{hexane}}}{P_{\text{hexane}} + P_{\text{heptane}}}$$

$$= \frac{10.25}{13.3}$$

$$= 0.771$$

$$X_{\text{hexane}}^{\text{vap}} = \frac{3.05}{13.3}$$

$$= 0.229$$

This answer could also have been worked out using the fact that the sum of the mole fractions in a mixture always adds up to 1.

It can be seen from this that the vapour is richer in the more volatile component (hexane) than the original mixture was – the more volatile component has a greater tendency to enter the vapour phase. This is important for the separation of liquids using fractional distillation.

Raoult's law can also be applied to dilute solutions containing small amounts of non-volatile solutes to calculate by how much the vapour pressure is lowered. The calculation is carried out in the same way assuming that the vapour pressure of the non-volatile solute is zero.

Worked example

D.19 The vapour pressure of water at 80 °C is 47.375 kPa. Calculate by how much the vapour pressure is lowered by dissolving 10.00 g of glucose (M_r 180.18) in 1.00 dm³ of water (density at 80 °C is 0.971 79 g cm⁻³).

$$\text{Mass of water} = 1000 \times 0.971\,79 = 971.79\text{ g}$$

$$\text{Number of moles of water} = \frac{971.79}{18.02} = 53.93\text{ mol}$$

$$\text{Number of moles of glucose} = \frac{10.00}{180.18} = 0.05550\text{ mol}$$

$$\text{Total number of moles} = 53.93 + 0.055\,50 = 53.98\text{ mol}$$

$$\text{Mole fraction of water} = \frac{53.93}{53.98} = 0.9990$$

$$\text{Vapour pressure of water} = 0.9990 \times 47.375 = 47.33\text{ kPa}$$

The vapour pressure of pure water was 47.375 kPa, therefore the vapour pressure decreases by 0.05 kPa when the glucose is dissolved.

The glucose is non-volatile and makes no contribution to the vapour pressure above the solution.

? Test yourself

- 14 For each of the following ideal solutions, calculate the vapour pressure of the solution and the mole fraction of each component in the vapour:
- a mole fraction of **A** = 0.400, vapour pressure of pure **A** = 20.0 kPa
mole fraction of **B** = 0.600, vapour pressure of pure **B** = 16.0 kPa
 - b number of moles of **C** = 1.20, vapour pressure of pure **C** = 10.0 kPa
number of moles of **D** = 0.800, vapour pressure of pure **D** = 7.00 kPa
 - c number of moles of **E** = 2.50, vapour pressure of pure **E** = 5.00 kPa
number of moles of **F** = 1.60, vapour pressure of pure **F** = 8.00 kPa
 - d mole fraction of **G** = 0.200, vapour pressure of pure **G** = 12.0 kPa
vapour pressure of pure **H** = 14.0 kPa
- 15 Calculate the decrease in the vapour pressure of water when 20.00 g of sucrose ($M_r = 342.34$) is dissolved in 500.0 cm³ of water at 40 °C. The vapour pressure and density of water at this temperature are 7.3812 kPa and 0.992 22 g cm⁻³ respectively.

The boiling point of a liquid

Boiling occurs when bubbles of vapour form in a liquid and escape. Bubbles of vapour cannot be formed until the vapour pressure equals the external pressure.

A liquid boils when its vapour pressure equals the external pressure. So the normal boiling point of a liquid is the temperature at which the vapour pressure of the liquid is one atmosphere.

The more volatile a liquid is, the higher its vapour pressure will be at a certain temperature – and therefore the lower its boiling point will be.

Dissolving a non-volatile solute in a solvent will increase the boiling point of the solvent (boiling point ‘elevation’) because the vapour pressure is lowered and therefore the liquid must be heated more so that its vapour pressure is equal to atmospheric pressure (see the calculation above).

Because the vapour pressure of a mixture of two liquids varies with the composition, the boiling point also varies with the composition of the mixture. The richer a mixture is in the more volatile component, the lower the boiling point.

Figure **D.47** is a boiling point–composition diagram for an ideal solution with two components – liquid A and liquid B. The red line shows how the boiling point changes with composition and the blue line shows the vapour with which a particular mixture is in equilibrium. A is the more volatile component (higher vapour pressure in Figure **D.46**) and has the lower boiling point.

Consider a mixture in which the mole fraction of A is 0.3 (and that of B is 0.7) – the green dashed line (in Figure **D.47**) is drawn up from this composition to the liquid line and it can be seen that the boiling point of this mixture is 81 °C. The horizontal line across to the vapour curve (black dashed line) indicates the composition of the vapour with which this liquid is in equilibrium. It can be seen that the mole fraction of A in the vapour is 0.6, compared to 0.3 in the liquid – the vapour is richer in the more volatile component.

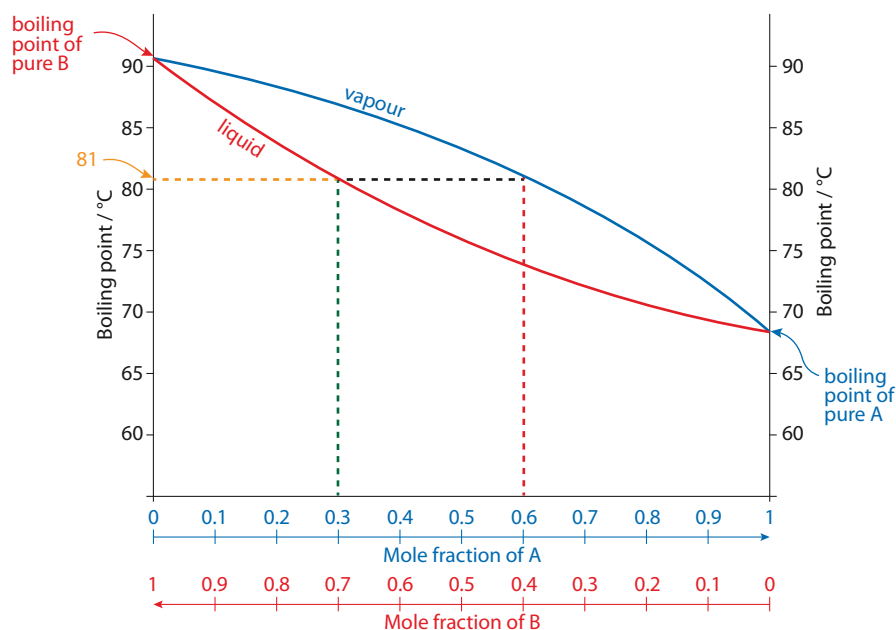


Figure D.47 A boiling point–composition diagram for an ideal solution.

Fractional distillation

The above discussion can be used to explain how fractional distillation separates a mixture of liquids.

If a mixture containing 0.1 mol A and 0.9 mol B is heated in the flask, it will boil at 87°C (Figure D.48). The mole fractions in the vapour will be $X_A = 0.3$ and $X_B = 0.7$ – so the vapour is richer in the more volatile component. The fractionating column is hot at the bottom and cooler at the top so, as the vapour rises up the column to the cooler parts, it condenses on the beads in the column to form a liquid with composition $X_A = 0.3$ and $X_B = 0.7$.

This liquid will then trickle down the column, where it is heated by hotter vapour coming up from below and will boil again, but this time at a lower temperature (81°C) because it is richer in the more volatile

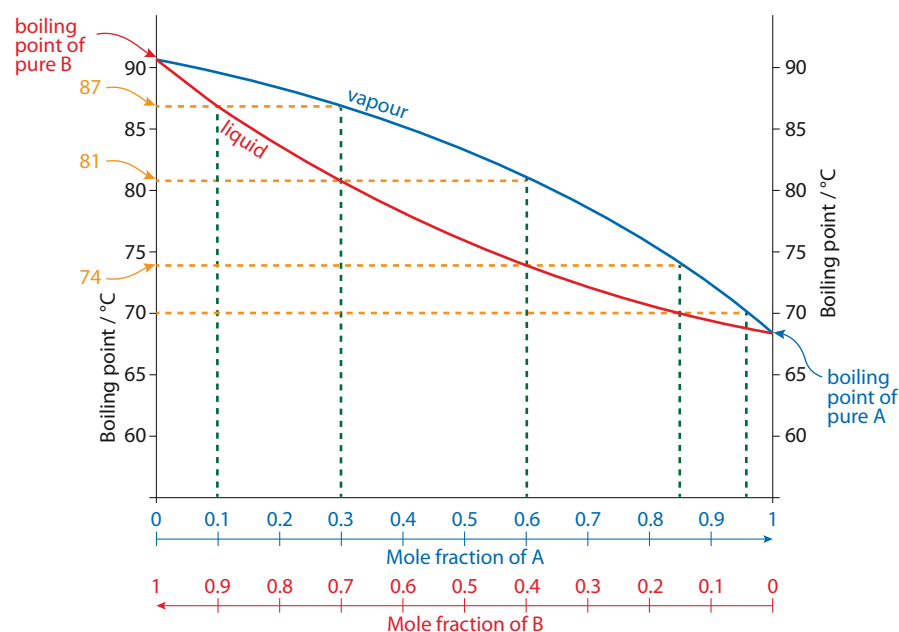


Figure D.48 Fractional distillation allows separation of a mixture of liquids.

component. This liquid will boil to form a vapour with composition $X_A = 0.6$ and $X_B = 0.4$. This vapour is richer in the more volatile component and rises higher up the column before it condenses to form a liquid of the same composition.

The liquid trickles down the column where it is heated again and boils at 74°C to form a vapour of composition $X_A = 0.84$ and $X_B = 0.16$. Each time the liquid boils and condenses it becomes richer in the more volatile component and rises higher up the column until, if the column is long enough, essentially pure A will be obtained from the top of the column.

The closer the boiling points of the liquids, the more series of boiling and condensing processes are required and therefore the longer the column required for good separation of the liquids.

Not all liquids can be completely separated by fractional distillation and, for instance, ethanol and water form a constant-boiling-point mixture (95.6% ethanol).

Spectroscopic identification of organic compounds

Once an organic compound has been synthesised, purified and extracted it can be identified using spectroscopic techniques.

The structure of many organic molecules can be determined using a combination of infrared spectroscopy, mass spectrometry and proton NMR.

The structures of pharmaceutical compounds are usually very complicated and their spectra are extremely difficult to interpret – here we look at some simpler examples.

Secondary amides

A secondary carboxamide (amide) group (RNH) is present in the structure of many drugs – oseltamivir and paclitaxel for example. The structure of a simple secondary amide is shown in Figure D.49.

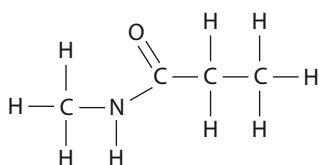


Figure D.49 The structure of a secondary amide; *N*-methylpropanamide.

The first step in identifying a compound is usually to run an infrared spectrum – this is a very simple and quick technique. The infrared spectrum of the secondary amide is shown in Figure D.50. The absorptions shaded red can be ignored – they are due to a different vibrational motion in the molecule (N–H bend). We look at the region above 1500 cm^{-1} to assign the bands in the spectrum:

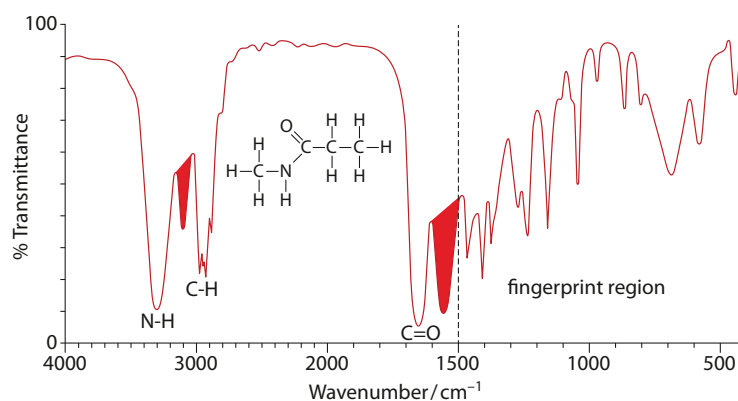


Figure D.50 The infrared spectrum of *N*-methylpropanamide.

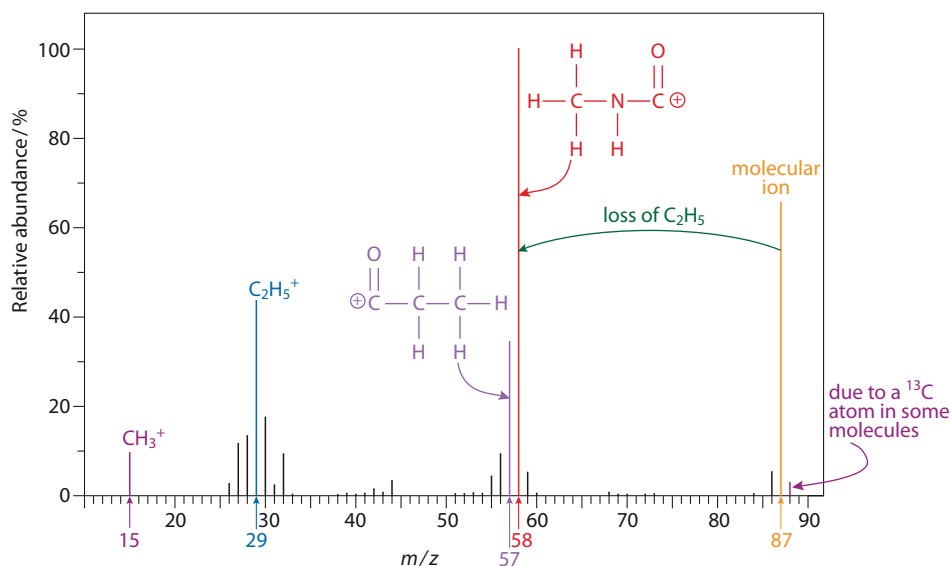


Figure D.51 The mass spectrum of *N*-methylpropanamide.

- the band at about 1650 cm^{-1} is due to $\text{C}=\text{O}$ – this is a bit lower than we normally expect for this bond but is fairly typical for amides
- the band at 3000 cm^{-1} is found in all organic spectra – it is due to the $\text{C}-\text{H}$ stretch
- the band at 3300 cm^{-1} is due to the $\text{N}-\text{H}$ bond.

Some key peaks are highlighted in the mass spectrum of the secondary amide in Figure D.51. The molecular ion peak occurs at 87 and this gives us the relative molecular mass of the molecule. The peak at 88 is due to the presence of a carbon-13 atom in some of the molecules. Loss of an ethyl group (mass 29) from the molecular ion results in the formation of the CH_3NHCO^+ ion ($m/z=58$), whereas loss of the CH_3NH group forms $\text{CH}_3\text{CH}_2\text{CO}^+$, at $m/z=57$. Some other fragments that are formed are shown in Figure D.51.

The high-resolution NMR spectrum of *N*-methylpropanamide is shown in Figure D.52. There are four different chemical environments for protons. Coupling between the protons in the ethyl group gives rise to a triplet (two H atoms on adjacent C) and a quartet (three H atoms on adjacent C). There is coupling between the H on the nitrogen and the Hs of the methyl group, resulting in the signal for the methyl Hs being split into a doublet (one H atom on adjacent N). However, the splitting of the signal due to the H on the nitrogen into a quartet (three H atoms on adjacent C) is not seen because of broadening due to other effects.

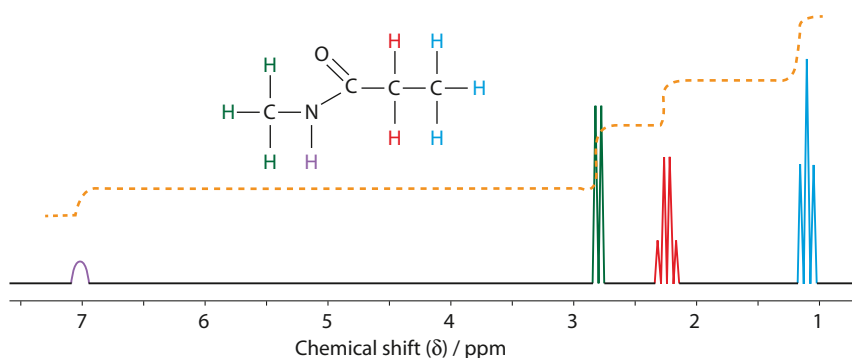


Figure D.52 The NMR spectrum of *N*-methylpropanamide.

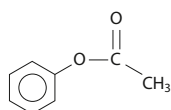


Figure D.53 The structure of phenyl ethanoate.

Phenyl ethanoate

Ester groups and benzene rings are components of the structure of many drugs – for instance, the structure shown in Figure D.53 is present in aspirin and diamorphine.

The infrared spectrum of phenyl ethanoate is shown in Figure D.54. The absorption at around 1750 cm^{-1} is due to the $\text{C}=\text{O}$ bond, but an ester also contains a $\text{C}-\text{O}$ single bond and the band for this is in the fingerprint region ($1050\text{--}1410\text{ cm}^{-1}$). The peaks at 1600 cm^{-1} and just below 1500 cm^{-1} are due to the vibrations of $\text{C}=\text{C}$ in the benzene ring – these bands are very characteristic of benzene rings.

The mass spectrum of phenyl ethanoate is shown in Figure D.55. Mass spectra can be difficult to interpret because of rearrangement reactions. For instance, in the mass spectrum of phenyl ethanoate, the large peak at $m/z=94$ is due to a rearrangement to produce $\text{C}_6\text{H}_5\text{OH}^+$. A characteristic peak in the mass spectrum of monosubstituted aromatic compounds is the C_6H_5^+ peak at $m/z=77$. Loss of CH_3CO from the molecular ion results in the peak at 93.

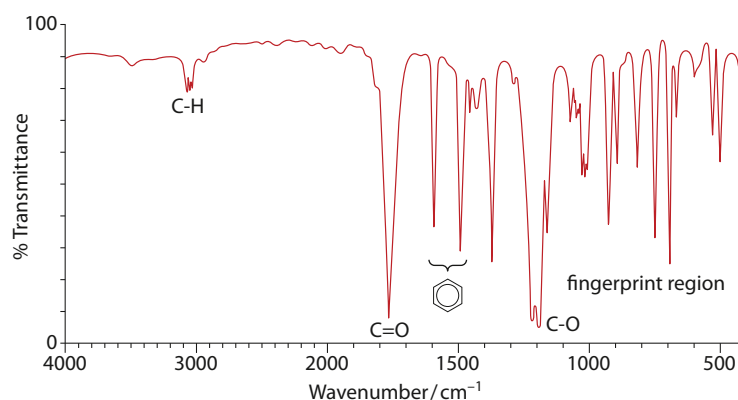


Figure D.54 The infrared spectrum of phenyl ethanoate.

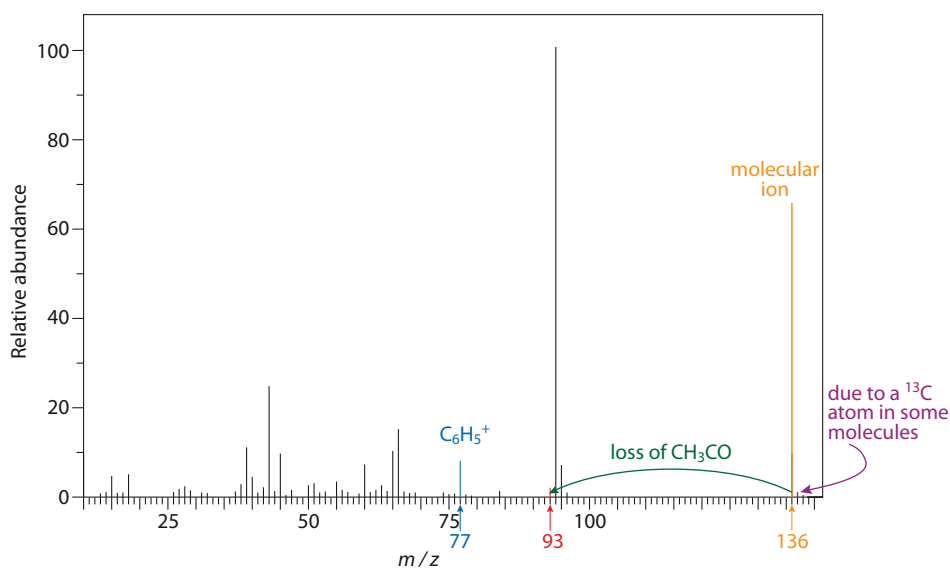


Figure D.55 The mass spectrum of phenyl ethanoate.

The NMR spectra of compounds containing benzene rings usually contain peaks around 7 ppm due to the hydrogen atoms attached to the benzene ring. These peaks may be resolved into separate signals or could just appear as one complex signal. An example of the NMR signal from the protons attached to a benzene ring is shown in Figure D.56.

Steroid detection in sport

Steroids are mostly non-polar molecules having a common structural feature known as the **steroid backbone** (Figure D.57). This is made up of three six-membered rings (labelled A, B and C) and a five-membered ring (D) fused together. Steroids vary depending on the type and position of substituents on this steroid backbone. There is also usually a carbon-carbon double bond in either ring A or ring B. Oestrogens are different to the other steroids in that they have an aromatic A ring.

Testosterone (Figure D.58) is an androgen (a male sex hormone). Androgens are also called **anabolic steroids** because they promote tissue growth, of muscle in particular. They are taken by sportsmen and sportswomen to enhance performance because they increase muscle mass and are also believed to improve endurance. The three most common anabolic steroids that are abused are testosterone and the synthetic derivatives nandrolone (Figure D.58) and stanozolol. They have been used in disciplines such as athletics, weightlifting and cycling.

The ethical implications of taking anabolic steroids are clear – it gives users an unfair advantage over their competitors and is simply cheating. It is a major concern to sporting bodies worldwide and drug screening in major sporting events is routinely employed to detect abuse.

Abusing anabolic steroids can have a major impact on the body – their use can cause a number of side effects such as breast growth in men, acne, infertility, mood swings and aggressiveness. They can also cause high blood pressure, liver disease (including cancer), heart attacks or strokes. Psychologically, abusers of anabolic steroids can become addicted to them, developing an increased desire to keep taking them, even if unwanted side effects occur.

Steroids and their metabolites (substances they are broken down into in the body) can be detected using the combined techniques of gas chromatography–mass spectrometry. The organic chemicals are extracted from a urine sample and then separated into their various components using gas chromatography (gas–liquid chromatography). Each band, as it comes out of the chromatography column, is passed directly into a mass spectrometer where it is analysed. The mass spectrum of a compound is generally characteristic of that compound and so, by comparison with the mass spectra in a database, the compound can be identified.

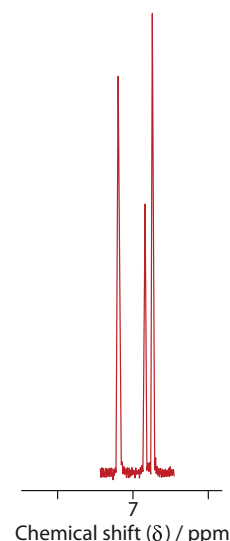


Figure D.56 Part of a typical NMR spectrum of an aromatic compound showing the signals due to the H atoms attached to the benzene ring.

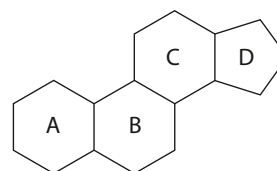


Figure D.57 Structure of the steroid backbone.

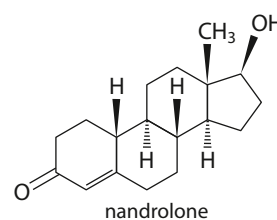
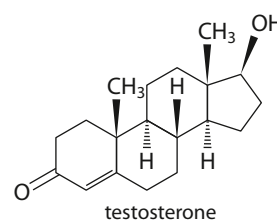


Figure D.58 The structures of testosterone and nandrolone.

Gas chromatography

Gas chromatography (GC) is used for the separation of mixtures, such as drugs in urine. Figure D.59 is a schematic diagram of a gas chromatograph. The sample is injected into a heated chamber, where it is vaporised.

- An inert gas (usually nitrogen or helium), called the carrier gas, carries the sample through the column.
- The column contains a non-volatile liquid (the stationary phase) spread onto the surface of finely divided solid particles. As the mixture is carried through the column, it separates into its components. Separation is by partition – so how quickly each component travels depends on its solubility in the liquid stationary phase. Those with higher solubility in the liquid stationary phase are slowed down in the column relative to those that are less soluble in the stationary phase. The **retention time**, the time between when the mixture is injected and the time it is detected, also depends (among other things) on how volatile the substance is – the most volatile compound generally has the greatest tendency to enter the gaseous phase and therefore the shortest retention time.
- Each component is fed into a mass spectrometer as it emerges from the column.

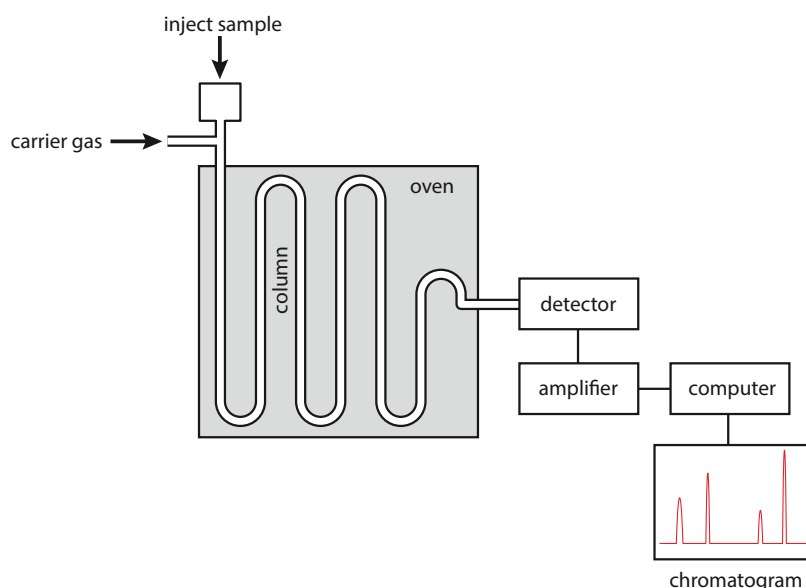
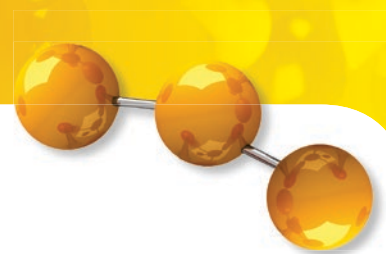


Figure D.59 A gas chromatograph.

Detection of alcohol in the breath

The presence of ethanol/alcohol in the breath can be detected using redox reactions – either using an oxidising agent in a chemical reaction or in a fuel cell.

A breathalyser test is a common test carried out at the roadside and involves the driver breathing into a device that detects the amount of alcohol in the breath. In the lungs, an equilibrium is established between alcohol dissolved in blood plasma and alcohol in the breath. Therefore, the amount of alcohol in the breath can be used to determine the amount of alcohol in the blood plasma.



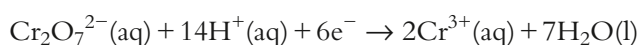
There are three main ways that alcohol can be measured in the breath.

The first uses a chemical test in a breathalyser, which either contains dichromate(VI) ($\text{Cr}_2\text{O}_7^{2-}$) ions in crystals or solution. Dichromate(VI) ions are orange and any ethanol present in the motorist's breath will cause a change in colour to green as they are reduced to chromium(III) ions – the ethanol is first oxidised to ethanal and then to ethanoic acid in the process. The degree of colour change is directly related to the level of alcohol in the breath. In the simple, disposable devices using potassium dichromate(VI) crystals, the amount of alcohol in the breath is determined by how many crystals change colour. In devices using potassium dichromate(VI) solution the amount of colour change can be measured using a colorimeter/photocell.

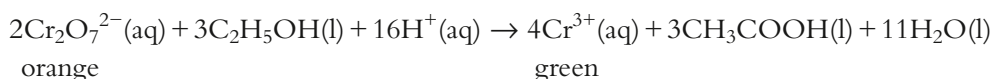
The equations for the chemical reactions involved are:



oxidation

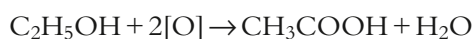


reduction



overall

The oxidation of ethanol to ethanoic acid may also be shown as:



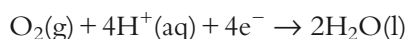
where [O] represents oxygen from the oxidising agent.

These chemical-based breathalysers have now largely been replaced by a newer, more accurate type of hand-held analyser that uses a **fuel cell** to detect and measure alcohol in the breath. In this type of device, the breath enters a fuel cell fitted with two platinum electrodes and any alcohol in the breath gets oxidised. Instead of the energy being released as heat, it is converted directly to electrical energy and an electric current is generated – the more alcohol present, the higher the current. The reactions that occur in the fuel cell are:



oxidation

occurs at the anode



reduction

occurs at the cathode



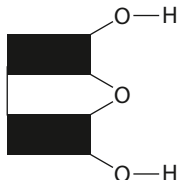
overall

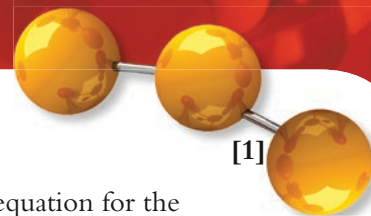
The third method for detecting alcohol in breath involves infrared spectroscopy.

Nature of science

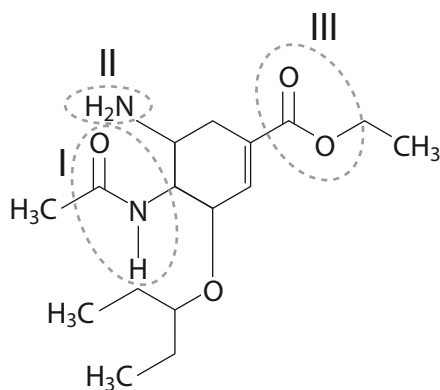
Taking drugs to enhance performance in sport is as old as sport itself. What has changed, however, is the ability of scientists to detect drugs. Advances in technology and better understanding of metabolic pathways have allowed the development of ever-more-sophisticated testing protocols for drugs in sport and their metabolites. But science is always improving and so blood/urine samples from the Olympics are kept for eight years so that future advances might be able to detect previously undetectable substances.

Exam-style questions

- 1 Drugs and medicines can have a number of physiological effects on the body.
- a Explain the meaning of the following terms:
 - i therapeutic effect [1]
 - ii side effect [1]
 - iii tolerance [1]
 - iv bioavailability. [1]
 - b Explain the term 'therapeutic window'. [1]
 - c Drugs can be administered to a patient via a number of different routes.
 - i There are three main ways of giving a drug by injection. Which of these gives the fastest response, and why? [2]
 - ii Which route would be chosen to treat a lung condition, such as asthma, locally? [1]
- 2 Analgesics and antibiotics are very important classes of drugs.
- a Describe how a mild analgesic, such as aspirin, causes an analgesic effect. [1]
 - b
 - i The structure of aspirin is shown in Figure D.5 and the IB Chemistry data booklet. Write an equation, showing structural formulas, for the synthesis of aspirin from salicylic acid. [2]
 - ii In an experiment to synthesise aspirin, 5.00 g of salicylic acid (M_r 138.13) was reacted with excess of the other reactant in the presence of a concentrated phosphoric acid catalyst. 5.02 g of a white solid was obtained at the end of the reaction. Calculate the yield of aspirin. [3]
 - iii Explain why it is possible to have a yield of aspirin greater than the initial mass of salicylic acid. [1]
 - iv State the name of the process by which aspirin can be purified and explain how it is carried out in the laboratory. [4]
 - c Briefly outline the mechanism by which penicillins carry out their antibacterial activity. [2]
 - d Some bacteria have developed resistance to penicillins by producing an enzyme that deactivates the penicillin.
 - i Name the enzyme produced by penicillin-resistant bacteria. [1]
 - ii Explain how this enzyme deactivates the penicillin. [1]
 - iii Which part of the penicillin structure can be modified to make the penicillin less sensitive to the actions of this enzyme? [1]
- 3 Analgesics are used to reduce pain.
- a Morphine and diamorphine are both strong analgesics. Describe how they carry out their analgesic effect. [1]
 - b The structures of morphine and diamorphine can be found in the IB Chemistry data booklet (and in Figure D.16).
 - i Describe how the structure of diamorphine differs from morphine, with respect to the functional groups present. [1]
 - ii State the type of reaction used to convert morphine to diamorphine. [1]
 - iii Write an equation for the reaction in part ii using the structure below to represent the structure of morphine. [2]
- 
- iv Describe two possible social problems that can occur through heroin addiction. [2]



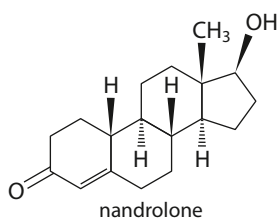
- 4 a Name the acid found in the gastric juice in the stomach. [1]
- b Calcium carbonate is an antacid used to neutralise excess acid in the stomach. Write the equation for the reaction between calcium carbonate and this acid. [1]
- c Explain the mode of action of ranitidine (Zantac®). [3]
- d Explain, using an example, what is meant by an 'active metabolite'. [2]
- e Buffers are important in some drug formulations. A buffer solution was made by dissolving a mixture containing 15.00 g of KH_2PO_4 and 15.00 g of K_2HPO_4 in water and making up to a total volume of 1.00 dm^3 . Given that the pK_a for H_2PO_4^- is 7.21, calculate the pH of the buffer. [3]
- 5 a How do viruses and bacteria differ in the way that they replicate? [2]
- b Describe **two** different ways in which antiviral drugs work. [2]
- c Oseltamivir (structure shown below) is an important drug used to treat influenza. Identify the functional groups that are highlighted in the structure. [3]



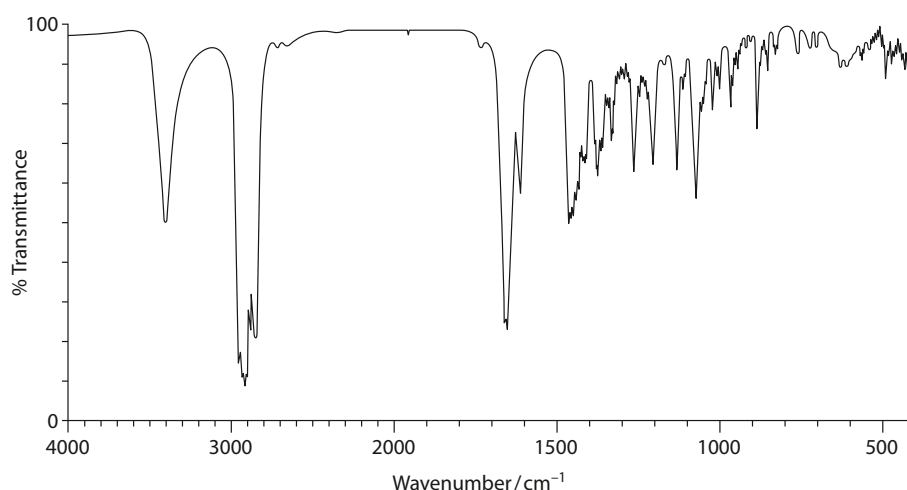
- d Explain why AIDS is so difficult to eradicate on a global scale. [2]
- 6 a Explain how antibiotics could enter the environment and a problem caused by this. [4]
- b Radioisotopes are used extensively in medicine. [2]
- i Distinguish between high-level nuclear waste and low-level nuclear waste. [2]
- ii State **two** sources of low-level nuclear waste in medicine. [2]
- iii Explain how low-level nuclear waste from medicine may be disposed of. [2]
- HL** 7 Taxol® is an important drug with several chiral centres. [1]
- a State which disease Taxol is used to treat. [1]
- b State the natural source of Taxol. [3]
- c The semi-synthesis of Taxol uses a chiral auxiliary. Describe the use of chiral auxiliaries in asymmetric synthesis. [3]
- d Describe how enantiomers can be identified by polarimetry. [3]

- HL 8 a** Lutetium-177 is a radioisotope that is used in medicine. Lutetium-177 undergoes β decay. [2]
- Write a nuclear equation for the decay of lutetium-177. [2]
 - A patient is given a 10.0 mg dose of lutetium-177. Given that the half-life of lutetium-177 is 6.71 days, calculate how much of the lutetium is present in the patient's body after four weeks. [4]
- b** State **three** side effects of radiotherapy. [3]
- c** Targeted alpha therapy is a relatively new form of radiotherapy. Explain how the technique works. [3]

- HL 9** After a drug has been synthesised it must be extracted from the reaction mixture and then purified. [3]
- a** In a procedure, a student poured the reaction mixture into a beaker containing 50 cm³ of ice-cold water and then carried out solvent extraction using three separate 50 cm³ portions of ethanol. Explain what is wrong with this procedure and suggest an improvement. [3]
- b i** Assuming that benzene and methylbenzene form an ideal solution, calculate the vapour pressure of a mixture containing 1.00 g of benzene and 10.00 g of methylbenzene at 20 °C. The vapour pressures of benzene and methylbenzene at 20 °C are 10.00 kPa and 2.93 kPa respectively. [4]
- ii** Explain how the composition of the vapour in part **i** compares to that of the original liquid. [2]
- iii** Explain the principles of a method for extracting benzene from the benzene–methylbenzene mixture in part **i**. [4]
- c** The structure of nandrolone, an anabolic steroid, is shown below:



- i** Identify three functional groups present in nandrolone. [3]
- ii** The infrared spectrum of nandrolone is shown below. Identify the absorption bands in the region above 1500 cm⁻¹ in this spectrum. [4]



- iii** Explain how nandrolone could be identified in a sample of urine from an athlete. [2]